Performance measurement of policy priorities: Tracking government performance

by

Charline Mouton

Dissertation presented for the degree of PhD in Evaluation Studies in the Faculty of Arts and Social Sciences at Stellenbosch University

Supervisor: Dr Lauren Wildschut

Co-supervisor: Prof Johann Mouton

March 2021

Declaration

By submitting this dissertation electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

March 2021

Abstract

There is universal consensus that governments need to execute their functions well so that the needs of citizens are met. Performance measurement enables governments to determine whether their efforts are effective. This thesis presents a focus on one approach to improving performance through rigorous attention to the delivery of public services – referred to as Deliverology by its originator, Sir Michael Barber. As head of the Prime Minister Delivery Unit (PMDU) in the Tony Blair Government in the United Kingdom (UK), Barber developed an approach that targets the implementation shortfalls in government. This thesis employs a case study approach to provide a practical example of how Deliverology was applied in the Western Cape Government (WCG) to implement and measure the performance of six strategic programmes, called Game Changers. The WCG under the auspices of Premier Helen Zille, established the Western Cape Delivery Support Unit (DSU) in 2015 to drive the delivery of the Game Changer programmes.

In executing the Deliverology approach, some of the perceived limitations in the performance measurement related elements of the approach were addressed through the inclusion of methods and features borrowed from main stream program evaluation (specifically theory-based approaches). The overarching aim of the study is to demonstrate that a modified Deliverology approach is an effective analytical framework to assess the performance of complex social interventions as represented by the eLearning Game Changer.

This study is divided into two Parts. Part 1 provides the historical roots of Deliverology, with a focus on the direct precursors as found within the performance measurement and policy implementation traditions. I utilise the three public sector regimes of the 20th century - Public Administration (PA), the New Public Management (NPM) and the New Public Governance (NPG) as a framework to show the theoretical and methodological advancements over time. Deliverology has roots in both the NPG and NPM and is put forward as a recent approach to solving government's implementation challenges

Part 2 covers the eLearning Game Changer case study, where I discuss how I have added to the Deliverology approach, drawing on the performance measurement and programme evaluation traditions. Deliverology's five-step process was utilised as the analytical framework for discussing the modified approach.

The gains from using an expanded approach were found to be three-fold: the explicit use of a theory-based approach elucidated the causal pathways; an additional sub step on indicator formulation contributed to greater clarity in conceptualisation and operationalisation of indicators; and finally, the introduction of a clear distinction between short and medium term outcomes mitigated the risk for an

unbalanced focus on outputs only. The distinction between short and medium term outcomes also assisted in setting realistic expectations as to what could be achieved within a relatively short period of time.

The study suggests that the modified approach is suitable for more complex interventions but requires the necessary technical capabilities and human resources to be put in place.

In conclusion, this case study demonstrates the value of a highly structured approach to performance measurement (as exemplified in the elements and routines and strategies of a modified Deliverology framework) when augmented with lessons learnt around theory-based monitoring and evaluation. Given the many policy reform and service delivery challenges in South Africa – basic health care, food security, education, inequality and many others – many of our social programmes are in fact complex interventions. This case study has argued for a very structured approach to tracking the performance and monitoring the outcomes of such complex programmes.

Opsomming

Dit word algemeen aanvaar dat regerings hul funksies goed moet uitvoer ten einde te voldoen aan die behoefte van landsburgers. Deur prestasiemeting kan regerings vasstel of hul pogings effektief is. Die studie dek ‘n benadering wat poog om regerings se prestasie te verbeter deur middel van ‘n sterk fokus op implementering – getiteld “Deliverology”. Barber het dié benadering ontwikkel toe hy aan die hoof was van die “Prime Minister Delivery Unit” in Tony Bair se kabinet. Met hierdie benadering wou Barber implementering verbeter ten einde te verseker dat meer intervensie-uitkomste bereik word. Ek volg ‘n gevallestudie benadering in die tesis, en verskaf ‘n praktiese voorbeeld van hoe Deliverology toegepas is in een van die Wes-Kaapse regering se ses strategiese programme (getiteld “Game Changers”). Premier Helen Zille het in 2015 ‘n Implementering-steuneenheid opgerig in die Wes-Kaapse regering ten einde oorsig te verskaf van die implementering van die ses “Game Changer” programme.

In die implementering van die Deliverology benadering, het sekere tekortkominge rondom prestasiemeting navore gekom wat gelei het tot wysings in die Deliverology benadering. Hierdie wysigings is gedoen binne die konteks van standaard program evaluering (spesifiek teorie-gebasseerde benadering). Die studie demonstreer hoe die aangepaste Deliverology benadering ‘n effektiewe raamwerk verskaf vir die prestasiemeting van komplekse sosiale programme, soos gevind word in die eLeer “Game Changer” program.

Die studie is opgedeel in twee dele. Deel 1 word gewy aan die voorlopers van Deliverology, met ‘n spesifieke fokus op die prestasiemeting, en beleidsimplementering tradisies. Ek gebruik die drie publieke sektor regimes van die 20ste eeu – naamlik “Public Administration” (PA), “New Public Management” (NPM) en “New Public Governance” (NPG) as ‘n raamwerk om die teoretiese en metodologiese bydraes van die tradisies uit te wys. Deliverology kan gekoppel word aan beide die NPG en die NPM, en word aangebied en bespreek as ‘n nuwe benadering wat kan help met regerings se implementeringsuitdagings.

Deel 2 beskryf die e-Leer “Game Changer” gevallestudie, waar ek die wysigings aan die Deliverology benadering bespreek. Ek gebruik die vyf stappe in die Deliverology benadering as ‘n raamwerk om die wysigings te bespreek.

Die studie maak drie bydraes: die eksplisiete gebruik van ‘n teorie-gebasseerde benadering tot program monitering verskaf meer duideliheid oor die kousale ‘pathways’; die invoeg van ‘n addisionele stap

rondom die formulering van indikatore wat meer duidelikheid gee wat die konseptualisering en operasionalisering van indikatore betref en laastens, ‘n duidelike onderskeid tussen kort en medium termyn uitkomste. Die onderskeiding tussen kort en medium termyn uitkomste verseker dat daar realistiese verwagtinge gestel kan word in wat bereik kan word met enige intervensie oor ‘n relatiewe kort periode.

Die studie suggereer dat die aangepaste benadering meer toepaslik is vir komplekse intervensies, wat vereis dat die toepaslike tegniese vaardighede en menslike hulpbronne in plek moet wees.

Ten slotte, die gevallestudie demonstreer die waarde van ‘n hoogs-gestruktureerde benadering tot prestasiemeting (soos geillustreer deur die elemente, roetines en strategiee van die aangepaste Deliverology raamwerk) en soos aangevul deur teorie-gebasseerde monitoring en evaluering. Daar is ‘n menigte beleidshervormingsinisiatiewe en dienslewering probleme in Suid-Afrika, byvoorbeeld basiese gesondheidsorg, voedselsekuriteit, onderwys, ongelykheid en baie ander. Baie van die sosiale programme wat onderneem word kan geklassifiseer word as komplekse intervensies. Hierdie gevallestudie demonstreer die waarde van ‘n hoogs-gestrukteerde benadering tot prestasiemeting in die geval van sodanige komplekse programme.

Acknowledgements

I am profoundly grateful to God, the Author of my life. You direct my steps and make plain my paths.

I would like to thank my supervisor and co-supervisor for their continued support and guidance. Your inputs kept me focused and provided direction when I needed it the most. I could not have done this without you.

I would like to express my sincere gratitude to my two former Western Cape Government DSU colleagues Jenny Cargill and Belinda Wood, and friend Liezel de Waal who took the time to read this study, and whose contributions added significant value. And then a huge thank you to the two data analysts in the DSU: Lance Quinn and Pauline van der Merwe who performed all the Game Changer data analysis – including the analysis included in this study; the long hours and hard work are greatly appreciated. Thank you also to the Western Cape Government, specifically the Department of the Premier and the Western Cape Department of Education for allowing me to undertake this study, and to utilise the data collected during the eLearning Game Changer period.

My biggest thanks go to my father, mother and two sisters. Your unwavering and loving support kept me going. Thank you Marike for assisting with the graphics, it made a huge difference. And finally, a special thank you to my late grand-mother, Charlotte Mouton who took a special interest in my studies and who could not wait for this work to be completed. This study is dedicated to you.

Contents

List of Tables

List of figures

List of Acronyms

3Es	Efficient, Economic and Effectiveness
AGSA	Auditor General of South Africa
ANC	African National Congress
ASER	Annual Status of Education Report
BoB	Bureau of the Budget
BSC	Balanced scorecard
CAPS	Curriculum and assessment policy statement
CEO	Chief Executive Officer
CeI	Centre for e-Innovation
CEMIS	Centralised Educational Management Information System
CFO	Chief Financial Officer
CLEAR	Centre for Learning on Evaluation and Results
CPRS	Central Policy Review Staff
CREST	Centre for Research on Evaluation, Science and Technology
DA	Democratic Alliance
DCOG	Department of Cooperative Governance
DOE	Department of Education
DfID	Department for International Development
DINOS	Delivery Units in name only
DoTP	Department of the Premier
DPME	Department of Performance Monitoring and Evaluation
DPSA	Department of Public Service and Administration
DSU	Delivery Support Unit
EBPM	Evidence-based policy making
EMIS	Education Management Information System
FMI	Financial Management Initiative

FMPPI	Framework for Managing Programme Performance Information
GAO	General Accounting Office
GASB	Governmental Accounting Standards Board
GNP	Gross National Product
GPRA	Government Performance and Result Act
GWM&E	Government-wide Monitoring & Evaluation
HoDs	Head of Departments
ICAI	Independent Commission on Aid Impact
ICT	Information and Communications Technology
IDP	Integrated Development Plan
IGF	Inter-Governmental Forum
IGR	Institute for Government Research
LAN	Local area network
M&E	Monitoring and Evaluation
MANCO	Management committee
MBO	Management by Objectives
MINMEC	Ministers and Members of Executive Councils Meeting
MOA	Memorandum of agreement
MOU	Memorandum of understanding
NDP	National Development Plan
NGO	Non-Governmental Organisation
NPO	Nonprofit Organisation
NPC	National Planning Commission
NPG	New Public Governance
NPM	New Public Management
NYBMR	New York Bureau of Municipal Research
OECD	Organisation for Economic Co-operation and Development

OMB	Office of Management and Budget
PA	Public Administration
PAR	Programme Analysis and Review
PART	Programme Assessment Rating Tool
PDIA	Problem Driven Iterative Adaptation Approach
PERSAL	Personal and Salary Administration System
PEMANDU	Performance Management and Delivery Unit
PESC	Public Expenditure Survey system
PFMA	Public Finance Management Act
PMDU	Prime Minister Delivery Unit
PPB	Planning Programming Budgeting
PPBS	Program, planning, and budgeting systems
PRC	Presidential Review Commission
PSA	Public Sector Agreement
RDU	Result and Delivery Unit
SA	South Africa
SASQAF	South African Statistical Quality Assessment Framework
SCI	Street crime initiative
TELI	Technology Enhanced Learning Initiative
TQM	Total Quality Management
TRDM	Targeting results, diagnosing the means
UK	United Kingdom
UNAIDS	United Nations Programme on HIV/AIDS
USA	United States of America
WAN	Wide Area Network
WCED	Western Cape Education Department
WCG	Western Cape Government

WPTS	White Paper on the Transformation of the Public Sector
ZBB	Zero Based Budgeting

Chapter 1: Introduction and rationale for the study

1.1 Autobiographical note

In 2014 the Western Cape Government (WCG) decided to adopt the Deliverology approach to measuring the performance (implementation and results) of six major interventions (called Game Changers). The Delivery Support Unit (DSU) – the organisational mechanism by which Deliverology is implemented – was subsequently established (2015) with the mandate to provide oversight and support the departments responsible for implementing the Game Changer programmes. Each Game Changer also had a DSU lead - and in some instances a data analyst - who worked closely with the implementing departments to support the delivery of the Game Changers. Performance tracking constituted a key part of the delivery system that was put in place and operated as a separate function.

As the Performance Tracking Director of the DSU, I was intimately involved in the performance measurement of most of the Game Changer programmes. I came to this position with a background in mainstream programme evaluation. I studied Monitoring and Evaluation (M&E) and completed a thesis on the history of programme evaluation in South Africa. As such, I was familiar with standard M&E concepts and principles but in the DSU, I faced a new challenge of measuring the performance of six government priority programmes, utilizing a "new" approach within the South African public sector to performance measurement.

My role could - from a methodological perspective – be described as a kind of participant observer: someone who was both participating in the implementation and monitoring of the Game Changers, but at the same time observing, recording and analysing how the Deliverology approach work in real life. In this role I had first-hand access to information that an external evaluator would not necessarily have, enabling me to identify the shortcomings but also the successes and gains of the approach. This thesis systematically documents my experiences in this regard.

1.2 Rationale for this study

My study centres around Deliverology - a specific and recent version of the "performance measurement" tradition that originated in the United Kingdom (UK) public sector in the early 2000s. Deliverology targets the implementation gap, positing that better implementation will lead to better results, and by implication, better government performance.

Public sector performance is a subject matter that transcends country borders, political constituencies, and bureaucratic arrangements. At the crux of governmental operations is the need to execute its functions well and to provide services that meet the needs of citizens. Understanding the role of

government and the effective execution of its functions has become increasingly important given the ever-changing eco-system within which government operates.

Some of the factors that require constant consideration by governments are globalisation, rapid technological advances, economic pressures, industrialisation, public views of the role and size of the state, rapid urbanisation and population growth (Hughes, 1998; Massey, 1993; Wessels, 2000, cited in Miller, 2005). Government reforms in response to this ever-changing environment, can either "span the entire public sector" or be "a surgical alteration" that addresses a particular aspect of government operations (Koma & Tshiyoyo, 2015, p.33). A performance imperative commonly underpins government reforms. The definition below demonstrates this, as it directly links the purpose of reform to improved performance.

> Public management reform is usually thought of as a means to an end, not an end in itself. To be more precise we should perhaps say that it is potentially a means to multiple ends. These include making savings (economies) in public expenditure, improving the quality of public services, making the operations of government more efficient and increasing the chances that policies which are chosen and implemented will be effective (Pollitt & Bouckaert, 2000, p. 6).

In South Africa the performance failings of the government are a constant refrain in the media. The need to tackle implementation is not new, as can be seen from these newspaper headlines:

- *Citizens tired of promises* (Saeed, 2019);
- *If SA wants to emulate Asia, implementation of NDP (National Development Plan) is key* (Modipa, 2015);
- *Policies need to be implemented* (Cronje, 2010);
- *Citizens negative about the future of SA* (Ndaliso, 2019);
- *SA policies lack implementation (*Pretoria News weekend, 2019);
- *South African business confidence at a 3-decade low in 2019* (Gulf Times, 2020).

Many reasons are cited as to why service delivery and the implementation of policies are often so challenging, especially within a developing country context. These include inadequately trained staff, ineffective intergovernmental and interdepartmental coordination, a lack of reliable data for making decisions, poorly framed policies, a lack of financial resources, a lack of project management skills and ineffective political and administrative leadership (Cameron & Tapscott, 2000; Koma & Tshiyoyo, 2015). In the absence of addressing these shortfalls, governments will continue to face the prevailing negative narrative around performance.

Although the South African government has made progress in addressing some of the inequalities of the past through the provision of access to basic services and payment of social grants, its performance still falls short, with mounting dissatisfaction with service delivery (Goldman, Mathe, Jacob, Hercules, Amisi, Buthelezi & Sadan, M., 2015).

Deliverology, as an approach to performance measurement and delivery, originated in the UK during Tony Blair's second term of office in 2001. Sir Michael Barber developed this approach during his time as the head of the Prime Minister Delivery Unit (PMDU) as a means of addressing government's implementation shortfalls. Following his departure from government, Barber continues to advise countries on implementation related reforms in his capacity as the founder and chairman of a consultancy called Delivery Associates. Almost two decades since the start of Deliverology it continues to receive widespread attention with many countries around the globe adopting the approach to improve government performance.

Given the widespread government reforms in South Africa, implemented after the first democratic election in 1994, it is not surprising that Deliverology would also eventually gain a foothold in the country. The influence of Deliverology is evident in the outcomes based approach instituted at national government level as well as the Operation Phakisa programme, but the first full manifestation of Deliverology occurred in 2015, when the DSU was established in the Office of the Premier in the Western Cape Government (WCG).

The backdrop of this study is the Western Cape, the fourth largest province in South Africa in terms of population and the 3rd largest when considering its economic contribution to the country. South Africa has a semi-federal government system which encompass three spheres of government: national, provincial and local (Goldman, Byamugisha, Gounou, Smith, Ntakumba, Lubanga, Sossou, Rot-Minstermann, 2018). Provincial governments have their own legislature and strong autonomy, in particular with regard to the developmental functions under their direct control, such as education, health, agriculture and social development (Goldman, Engela, Akhalwaya, Gasa, Leon, Mohamed & Phillips).

During her time as the Premier of the Western Cape province, Helen Zille, inspired by the work of the PMDU in the UK, saw the potential of Deliverology to address the service delivery shortfalls in the province.

The Western Cape is nationally recognised as the best performing province by the Auditor General of South Africa's (AGSA) findings. With this solid base of good governance established during Premier

Zille's first term of office, she wanted to shift the focus to the delivery of tangible results as expressed in her 2014 State of the Province address:

> It has become increasingly evident that we need a new way of delivery, one that is uncompromisingly performance driven and results-focused. (WCG, 2014)

This led to the establishment of the Western Cape DSU in 2015, with the aim of supporting the delivery of six cabinet-approved strategic priorities, called "Game Changers". Seven broad priority policy areas (Game Changers) were identified, with six falling under the purview of the DSU. These Game Changers were advancing apprenticeships for priority economic sector, quality after school programmes of the most disadvantaged learners, energy security, the better living model (a mixed use, mixed income property development), alcohol harms reduction and eLearning.

Given the relative novelty of Deliverology, this would always be an "experiment" within the South African context – a process of trial and error. The aim of the thesis is therefore to report on the results of this "experiment": to document how the Deliverology approach was followed in measuring the performance of one specific intervention – the eLearning Game Changer - and whether the "experiment" can be assessed to have been successful.

1.3 Research aims and objectives of this study

The main aim of my study can be formulated as follows: To demonstrate that a modified Deliverology approach is an effective analytical framework to assess the performance of complex social interventions – viz. Game Changers. In order to address this research aim, the study is 'disaggregated' into two subsidiary research objectives: First, to describe the key features of the Deliverology approach as it has evolved elsewhere in the world; and second, to describe and reflect on how this approach was modified in the Western Cape and how these modifications – which were informed by key elements from mainstream programme evaluation traditions – produced clear and demonstrable gains in assessing the performance of the eLearning Game Changer (the case selected for this study).

The modifications to the Deliverology approach in the Western Cape were applied to four of the five Game Changers but the eLearning Game Changer was selected as the case for this study for two reasons: first, the eLearning Game Changer was by far the most systemic of all the Game Changers. The budget for the eLearning programme over the three year Game Changer period was substantial and is estimated

at R1.1 billion (approximately US$73 million)[1]. In terms of scale, the eLearning programme was to be rolled out in ±1500 public, ordinary schools which translates into more than 30 000 teachers and approximately 1 million learners being reached. Secondly, the eLearning Game Changer addressed issues that are crucial to the digital age and how to produce teaching and learning through online and blended modes of delivery. The importance of this has been reinforced by the Covid-19 pandemic.

The relationship between the overall aim of the study and the subsidiary research objectives are illustrated in the figure below.

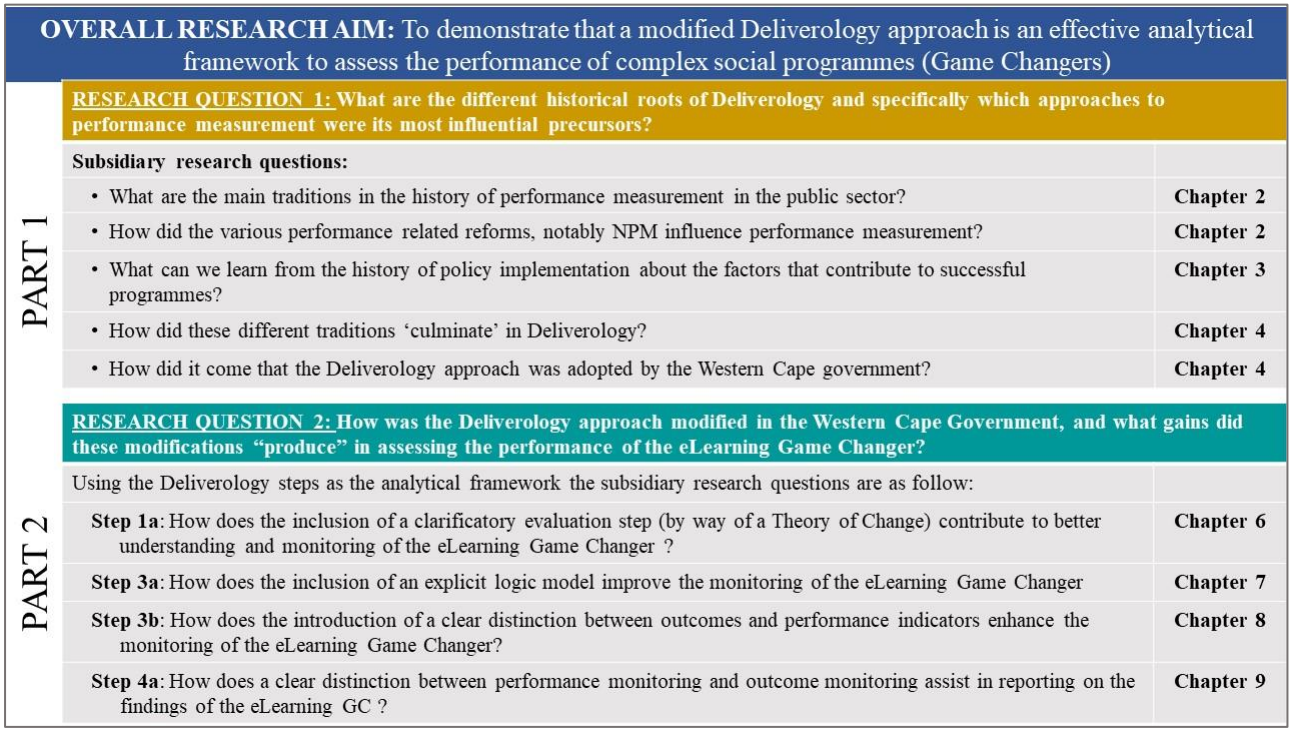


Figure 1: Main research aim and subsidiary research objectives of the study

Part One of the thesis is devoted to the first research objective and addresses the historical roots of Deliverology and specifically which approaches and traditions in performance measurement and policy implementation can be regarded as its direct precursors. This discussion is wide-ranging and include reconstructing the influence of the history of performance measurement, the role of NPM as well as the

[1] Amount obtained from the then head of the DSU. The R1.1 billion includes Western Cape Education Department (WCED) budget over three years as well as the proportion of broadband investment into schools. A further R200 million was provided for eLearning programme (beyond the Game Changer period) for 2019/2020 financial year which is not included in the R1.1 billion.

history of policy implementation. It also discusses the more immediate origins of Deliverology in the UK in the 1990's and its subsequent expansion to other parts of the world. And finally, I describe the origin, evolution and 'institutionalisation' of Deliverology in the Western Cape DSU. The key research questions addressed in Part One are:

- What are the main traditions in the history of performance measurement in the public sector?
- How did the various performance related reforms, notably NPM influence performance measurement?
- What can we learn from the history of policy implementation about the factors that contribute to successful programmes?
- How did these different traditions 'culminate' in Deliverology?
- How did it come that the Deliverology approach was adopted by the Western Cape government?

Part One concludes with a summary discussion of the Deliverology framework which forms the analytical framework for Part 2 of the study.

Part Two is devoted to a discussion of the second and arguably more important research objective of the study: to describe and reflect on how Deliverology was modified in the Western Cape and how these modifications – which were informed by my reading of key elements from mainstream programme evaluation traditions – produced clear and demonstrable gains in assessing the performance of the eLearning Game Changer.

In Part two I discuss in each chapter the gaps and shortcomings of the existing version of Deliverology, followed by an extensive discussion of the main changes and modifications that were made in the course of the implementation and monitoring of the eLearning game changer. This discussion is organized around the five steps typically included in Deliverology. Three of these steps are of particular relevance to my study (Step 1, Step 3 and Step 4 in light blue – Figure 2):

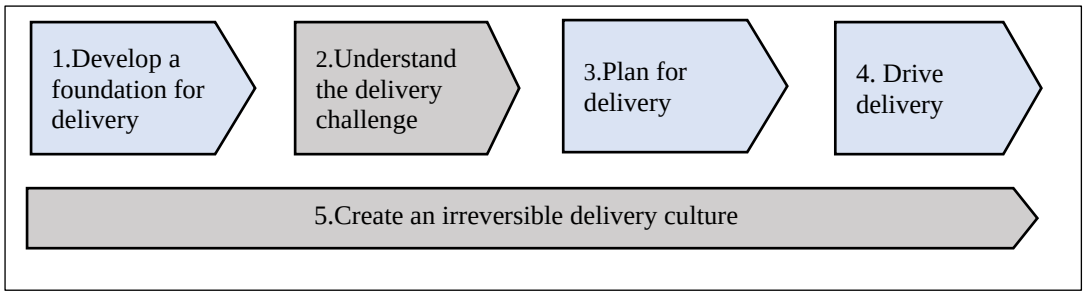


Figure 2: Deliverology framework

The main changes that were introduced to the Deliverology approach are:

- Step 1: The inclusion of an explicit clarificatory evaluation step in the framework and hence an emphasis on the necessity of formulating an explicit theory of change for the eLearning Game Changer as well as drawing a clear distinction between outputs, short-term and medium-term outcomes;
- Step 3: Much more attention given to the formulation of performance and outcome indicators to allow for the collection of granular and rich data; and
- Step 4: Clear separation between output monitoring and outcome monitoring (short-term and medium-term outcomes) which allows for more specific and to the point analysis and understandings of the findings of the eLearning Game Changer.

This translates into the following subsidiary research questions for Part 2 (Figure 1):

- How does the inclusion of a clarificatory evaluation step (by way of a Theory of Change) contribute to better understanding and monitoring of the eLearning Game Changer (Step 1a)?
- How does the inclusion of an explicit logic model improve the monitoring of the eLearning Game Changer (Step 3a)?
- How does the introduction of a clear distinction between outcomes and performance indicators enhance the monitoring of the eLearning Game Changer (Step 3b)?
- How does a clear distinction between performance monitoring and outcome monitoring assist in reporting on the findings of the eLearning GC (Step 4a)?

The chapters are organised as follow:

Chapter 2 presents an overview of the evolving thinking around the notion of performance, as well as the main paradigms and reforms associated with performance measurement in the public sector. The emphasis is on the USA and UK, as these two countries have been the forerunners in developing and

implementing performance-related reforms. I conclude Chapter 2 with an overview of performance measurement in South Africa, and how it is organised within the Government-wide Monitoring & Evaluation (GWM&E) system.

Chapter 3 covers the different "generations" of scholarships of implementation research as these evolved under the PA as well as policy implementation under the more recent paradigms of NPM and NPG.

In Chapter 4, I discuss the UK-originated approach of Deliverology. This encompasses the underlying principles of the approach, its achievements and criticisms, as well as the steps in executing the approach. I also introduce the Western Cape DSU and reflect on the institutionalisation of Deliverology as found in the DSU.

Part 2 is devoted to a detailed discussion of our case study: the eLearning Game Changer.

In Chapter 5 I introduce the eLearning Game Changer and elaborate on the design and methodological choices that informed our monitoring of the performance of the Game Changer between 2015 and 2019.

In the subsequent chapters (chapters 6 to 9), I show how I expanded and modified the Deliverology approach over time by a) introducing additional performance measurement activities and b) applying theory-based programme evaluation to it.

I discuss the main findings and some recommendations in the concluding chapter (Chapter 10).

1.4 Contribution of the study

There are at least two ways in which it is hoped this study would contribute to the body of knowledge.

Firstly, government performance is a topical issue that has evolved over time to be results driven, placing the needs of the citizen at the centre. Deliverology, as a means of addressing results by rectifying the implementation shortfall, was fully applied in South Africa for the first time through the establishment of the DSU in 2015. Although case studies have been done to cover Deliverology within

a developing country context, these studies tend to emphasise the Deliverology success factors and only provide a summary of the targets achieved. Through this study, the researcher provides a detailed and critical reflection of the application of Deliverology in a provincial government context.

Secondly, Deliverology, performance measurement and programme evaluation all share a similar purpose of achieving performance. However, Deliverology explicitly distances itself from standard programme evaluation, calling for "nimble" performance measurement to be undertaken. The programme evaluation and performance measurement culture that have been developed in South Africa over the past twenty years cannot be discounted as having no relevance to Deliverology. Similarly, it is worth considering how the philosophy and building-blocks of Deliverology can add value to existing approaches to performance measurement. The selected case, the eLearning Game Changer, combines these approaches; something which has not been undertaken before.

PART ONE

INTRODUCTION TO PART ONE: Osborne's three regime classification

As indicated in the introduction part one of this thesis is devoted to a detailed historical discussion of the performance measurement tradition (Chapter 2), the policy implementation research tradition (Chapter 3), and Deliverology (Chapter 4) as a relatively new tradition which has features in common with both. My focus is on the USA and the UK given that these two countries have been the forerunners in establishing and advancing these two traditions. The UK is also important as Deliverology originated there in the 2000's.

My historical review of these different traditions is informed and couched in terms of the periodisation of public sector 'regimes', as expounded by Osborne (2010) and others (Runya, Qigui & Wei, 2015). According to Osborne (2010), three major public sector regimes have been in operation since the late nineteenth century: i) Public Administration (PA) of the late nineteenth century through to the early 1980s, ii) New Public Management (NPM) of the 1980s through to the start of 21st century, and iii) the New Public Governance (NPG), which is the prevailing regime (Osborne, 2010). The PA is characterised by a clear divide between politics and administration, a focus on the rule of law and the centrality of the bureaucracy in policy formation and policy implementation (Osborne, 2010, p. 2). The NPM introduced private sector and managerial principles into the public sector.

On the policy implementation front this led to contracting out of services as well as emphasis being placed on performance management (Osborne, 2010). The NPG takes a broadened perspective on policy implementation rooted in institutional theory and network theory. In essence, the NPG suggests that policy implementation is undertaken by multiple actors (inside and outside of government) which brings into play many other institutional and external variables (Osborne, 2010). Two comments warrant mention around the use of this periodisation: Osborne (2010) acknowledges that this categorisation can be viewed as an oversimplification as regimes tend to overlap and co-exist. Also, at the time of suggesting this broad categorisation, it was still being debated whether NPG does in fact constitute a new regime. Regardless of this, it provides an overall framework that spans all the traditions covered in Part 1.

The placement of Deliverology in the figure below reflects (a) the fact that it is firmly located in the NPG; as well as (b) that it has 'borrowed' elements of the performance measurement and policy implementation traditions.

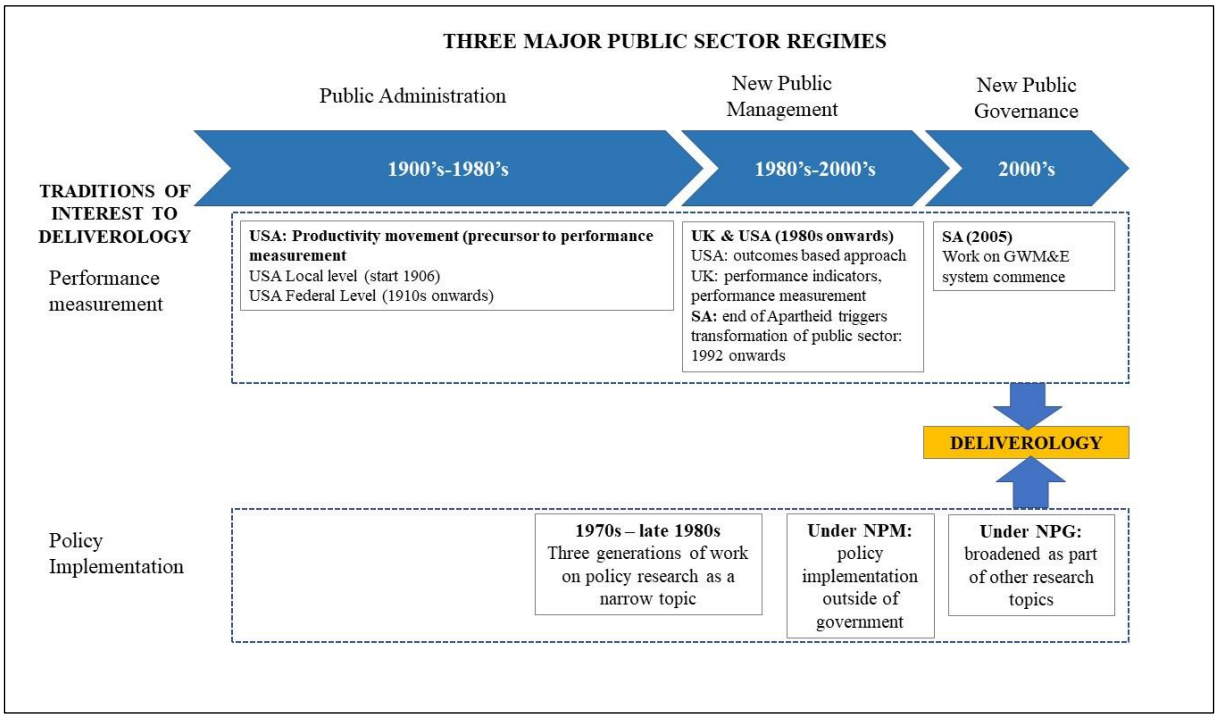


Figure 3: Historical overview of the evolution of different traditions pertaining to the performance of public sector programmes

Figure 3 shows how the performance tradition can be mapped onto the three public sector regimes as identified by Osborne. Early efforts related to performance in the public sector took place under the banner of the productivity movement, which started at the beginning of the 1900s. Federal government productivity efforts in the USA followed soon thereafter but were characterised by greater fluctuation as the prevailing socio-economic and political context determined the rationale and intensity of performance related reforms. Only in the 1970s did the concept of performance come to be associated with efficiency, effectiveness and economy. This development was directly linked to reforms initiated under the NPM movement in the 1980s and 1990s in the USA and UK.

It is worth pointing out that the NPM also influenced the South African post-Apartheid government (1994) thinking at the time. Examples of this include the introduction of decentralisation (giving managerial responsibilities to managers and delegating powers to provinces), corporatisation (converting departments into public entities) and downsizing or rightsizing (reducing the size of

government) (Cameron, 2009). Another direct consequence of NPM was the increased internal demand for performance information as a means of exercising control over the many stakeholders involved in executing government functions. The centrality of the citizen, and the need to demonstrate value for money, as well as being accountable for results, also added external pressure for performance information. In 2005, Cabinet approved the development of a GWM&E system, locating this work under the NPG regime as per Figure 3.

From the NPM onwards the terminology associated with performance in the public sector expanded considerably. Over the past twenty years and more, terms such as "performance measurement", and "performance indicators" have become strongly embedded in how governments across the world approach their responsibilities to monitor and account for goods and services delivered with public funds. The term "performance indicators", which is now a pervasive word in many government performance reports, is linked to the Financial Management Initiative (FMI) in the UK, with the intent that these indicators should measure the spectrum of government operations, i.e. inputs, outputs as well as outcomes. This correlates with Wholey and Hatry's (1992) description of what performance monitoring entails: "…They [performance monitoring systems] go beyond the more typical measurements of program costs, services delivered, and numbers served. Performance monitoring typically covers short-term and medium-term outcomes of program activities" (Wholey & Hatry, 1992, p.605).

The evolution of the policy implementation and implementation research tradition is displayed in Figure 3 as well. In the early 1970s, policy implementers started acknowledging the interrelatedness between implementation and outcomes and how poor implementation often leads to the non-achievement of outcomes. In the policy sciences, three generations of policy implementation research can be distinguished (from the early 1970s to mid-1990s) (Goggin, 1986; Hupe & Sætren, 2015; Schofield, 2001). Located mainly within the PA regime, the three generations of scholars attempted to construct a theory of implementation that would explicate the success variables of implementation. The first generation scholars (1970s) subscribed to a rational, linear approach to theory building, which considered very few variables and were mainly qualitative in nature. The second generation scholars (divided between top-down and bottom-up scholars) advanced the field by developing analytical frameworks to empirically test the variables that influence implementation. The third generation scholars abandoned attempts at developing a single theory for policy implementation. Instead their focus was on synthesising the many success variables by undertaking more quantitative research and comparative studies.

With the onset of the NPM, interest in policy implementation waned as private sector principles were introduced in the public sector, leading to much of policy implementation increasingly being located outside of government. The main underpinning of the NPG, as a response to the NPM, is that effective policy development and implementation is reliant on stakeholders, inside and outside government, working together to achieve the policy objectives. The greater emphasis on the cross-cutting nature of implementation as well as returning control to the centre of government are just two of the key characteristics of the NPG.

Deliverology[2] is a recent reform that aims to rectify the imbalance between policy formation and policy implementation - to achieve results, significantly more effort should go towards policy implementation. The Deliverology approach has evolved over time: not only has the author of this approach (Sir Michael Barber) refined and adjusted the framework, but countries instituting the approach have adjusted the approach to fit their context. This was also the case in the WCG, where the DSU adjusted and expanded Deliverology drawing on different paradigms.

Deliverology, located at the centre of Figure 3, has strong roots in the performance measurement tradition, particularly the NPM as the approach hinges on the availability and use of data. In terms of policy implementation, there are many synergies between the three generations of policy implementation and Deliverology, which leads one to agree with other scholars that not all of Deliverology is new (Birch & Jacob, 2019; Schacter, 2016). It is also of significance that Deliverology commenced under the NPG regime. This regime is associated with a greater focus on the horizontal dimension of implementation, as well as returning control to the centre of government post NPM. Delivery units – the organisational mechanism for Deliverology are typically established at the centre of government and are focused on the cross-cutting policy priorities.

[2] As per Manning and Watkins (2013), Deliverology is referred to here as an approach as opposed to a methodology. A methodology suggests that certain techniques and tools have been tested and validated to achieve a particular objective. Even though tools and techniques are available as part of Deliverology, these tools and techniques have not been validated as the means of solving delivery challenges and hence the decision to refer to this as an approach. But more importantly, Deliverology is better understood as a more comprehensive 'approach' to performance measurement, as it also includes core assumptions about the institutionalisation of delivery systems.

Chapter 2: Performance measurement in the public sector

2.1 Introduction

A quote from Downs and Larkey (1986, p. 59) sets the scene for this chapter on the history of performance measurement in the public sector:

> Since the beginning of the century, the development of quantitative, summary measures[3] of performance have been a centrepiece of most attempts to improve governmental performance. They lie at the core of the productivity movement in general and specific reform attempts such as program, planning, and budgeting systems (PPBS) and Management by Objectives (MBO). Not surprisingly, the inspiration behind this approach is heavily rooted in the business method folklore … It tells us that in a well-run company, managers are constantly informed and inspired by performance data. Many of the public sector's problems, it is assumed, stem from the absence of comparable information in public bureaucracies.

Five points from this quote that will be examined in this chapter:

a) Performance measurement has a long history that can be traced back to the beginning of the 1900s;
b) The purpose of performance measurement is to improve the performance of government;
c) Performance measurement is about quantification and making performance measurable;
d) Performance measurement is viewed as a panacea for many of governments' problems;
e) Multiple reforms have been launched over the years to improve government performance, many of these imported from the private sector.

Utilising Osborne's regime classification, a historical account will be provided of performance measurement which covers the USA, UK, as well as the South African public sector.

This chapter would not be complete without a discussion of the shortcomings associated with performance measurement, as well as the underlying causes of these shortcomings. Two of these

[3] In the early performance measurement history, the terminology "measures" were commonly used. Later in this chapter we show the shift in terminology to indicators under the FMI, and beyond. We also elaborate on the preferred use of indicators as opposed to measures in Chapter 8

underlying causes will receive greater focus, namely the difficulties with setting objectives and outcome measurement in a government environment.

2.2 The players that shaped performance measurement during the PA regime

The PA regime spanned most of the twentieth century and is associated with the following:

- The dominance of the "rule of law";
- A focus on administering set rules and guidelines;
- A central role for the bureaucracy in making and implementing policy;
- The split between politics and administration within public organisations.

(Hood, 1991, cited in Osborne, 2010, p. 2)

The guiding principles of the PA set the scene for both this chapter and the next chapter, where I will provide an overview of policy implementation. In this chapter the emphasis is on the measurement of government activities, covering the key players and advancements made during the three political regimes. The clear divide between politics and administration during the PA played a significant role in the way measurement activities were approached during this period. Other factors that shaped performance measurement activities included the prevailing economic and political landscape which found expression in the key concerns of government at specific points in time. The first eight decades of performance activities are characterised by waves of interest in efficiency, economy, productivity and ultimately quality. Bouckaert (1990) places the work done at municipal, state and federal level on efficiency, effectiveness and economy under the banner of the productivity movement (Bouckaert, 1990). I first cover a chronological overview of productivity efforts at local government and federal level, where after some definitional aspects are covered.

2.2.1 Local government measurement activities

Measurement activities started mainly in the local government arena in the USA during the latter part of the 1800s. This was motivated by the need for better government following widespread corruption at local level due to the Jeffersonian government not paying sufficient attention to the legislative control of local government (Williams, 2004). Greater demands were placed on governments to perform: "A generation ago a municipal government was considered commendable if it was honest. Today we demand a great deal more of our public service. It must be not only honest but efficient as well" (Ridley

& Simon, 1943, p. 1). The premise was that a clear separation of power between the administration and politicians is needed for government to operate efficiently.

The National Municipal League was subsequently established in 1894 as a coordinating body to align city reform efforts and rectify the effects of the Jeffersonian government (Ridley & Simon, 1943). Performance measurement – though not recognised as such at that stage offered a way to rebuild the profile of government through improved efficiency and greater transparency (Williams, 2004).

In 1906, the Bureau of City Betterment was established with the aim of "promoting the applied study of public administration in its formative years" (Williams, 2003, p. 6). In 1907, the agency changed its name to the New York Bureau of Municipal Research (NYBMR), with the objectives of making available empirical data that can capacitate government, support decision-making, as well as support cost and accounting practices reform efforts (Williams, 2004). These objectives were influenced by the three directors' prior involvement (William Allen, Henry Bruere and Frederick Cleveland) in social welfare agencies and settlement houses. As a result, surveys were frequently undertaken of entire communities and specific topics (e.g. housing), and to investigate local government conditions. Initially these surveys were mainly qualitative in nature, but this changed in the 1920s when the quantification of social data became advanced through the work of academics such as Karl Pearson and Ronald Fisher (Williams, 2004). Also linked to the general quantification efforts of social data, the NYMBR already in 1910 introduced a point system to standardise survey results. This resulted in the development of scorecards, which is viewed as the forerunner of index construction (Williams, 2004). Under the auspices of Frederick Cleveland, significant work was done on the budgeting and accounting front (Williams, 2004). The NYBMR believed that greater financial control would be obtained by way of a function-orientated budget. Additionally, functional categories would assist in determining efficiency and hence the bureau set out to develop functional categories which allowed for funding to be linked to work units (Ridley & Simon, 1943).

Another influencing factor of the work of the NYBMR was the scientific tradition. Already in 1887, Woodrow Wilson suggested a more scientific approach to public administration (Heinrich, 2003). This entrenched a rational decision-making approach to the undertakings of the bureau (Lynch & Day, 1996, Williams, 2002).

> The bureau and early public administration theorists largely assumed that the appropriate approach to effective decision-making was rationality: define one's objectives, define alternatives to meet those objectives, and select the proper course of action (Lindbloom, 1959, cited in Lynch & Day, 1996, p. 406).

The scientific tradition started in the field of mechanical engineering where a more scientific approach to manufacturing production was being sought (Ridley & Simon, 1943). Frederick Taylor, the father of the scientific management tradition set out to develop objective measures of production processes through observation and measurement, as he deemed this as the only way to improve work processes (Lynch & Day, 1996). The emphasis was on the individual worker, linking financial incentives to increased efficiencies, as opposed to organisational effectiveness which fell within the domain of financial management (Radnor & Barnes, 2007).

Drawing on the scientific management tradition, a variety of human resource practices were instituted to ensure officials were operating optimally, for example, record-keeping practices were introduced to track work outputs (Williams, 2003). Not only were the best ways of doing the job specified, but standards were also set in terms of time and resources to be utilised (Williams, 2004).

Outside of the NYBMR, other local government measurement-related initiatives included the development of municipal effectiveness measures by the National Commission on Municipal Standards (1928), as well as a handbook that aimed to assist city managers in assessing their performance quantitatively, issued by the International City Management Association in 1933 (Downs & Larkey, 1986, p. 66). The American Political Science Association also contributed towards advancing measurement activities within the municipal space: two national conferences on the Science of Politics were organised, respectively in September 1923 and 1924 (Ridley & Simon, 1943). At the first annual meeting in Madison, Wisconsin (1923), Lent Upson led a session around developing standards for municipal activities (Ridley & Simon, 1943). At the second annual meeting (1924), the development of a method to rate the efficiency of cities was discussed (Ridley & Simon, 1943).

2.2.2 Federal government measurement activities in the USA

Although local government took a clear lead, federal level efforts around productivity and efficiency did not lag far behind. However, the federal level can be characterised as being more sporadic. This is attributed to the fact that different factors motivated the interest in efficiency and productivity in the federal government. Early federal efforts (pre 1940) were mainly driven by the need to improve the performance of government, with efficiency viewed as the means of achieving this goal. Commitment to the efficiency agenda came through the establishment of the Commission on Economy and Efficiency established in 1913 under the Taft Administration, as well as the Institute for Government Research in 1916 (Bouckaert, 1990). The emphasis on efficiency with the purpose of improving the performance

of government continued under President Roosevelt's term (Bouckaert, 1990). In 1937 the Brownlow Committee, issued its report on "The efficiency of government", cautioning against a superficial application of efficiency; and suggesting instead that efficiency should permeate the full machinery of government (Bouckaert, 1990). According to Hubbell and Kinghorn (1988), the first significant federal level productivity effort came by way of the establishment of the Advisory Committee on Management Improvement, as well as the signing of Executive Order 10072 by President Truman in 1949. The Executive Order prescribed the frequent appraisal of government activities, while the Advisory Committee ensured the availability of funding to support efficiency studies (Hubbell & Kinghorn, 1988).

World War II and the onset of the Great Depression led to a cost cutting environment, resulting in various budgetary reform initiatives. In addition to efficiency, economy and productivity became a key concern of government in the USA between the 1940s and 1970s (Bouckaert & Halligan, 2006). At its core, budgeting serves three purposes: planning, management and control (Shick, 1966, cited in McNab & Melese, 2003). The earliest attempts at instituting budgeting procedures can be traced back to the civil war and the introduction of a tax levy: "Departments prepared detailed requests for spending and submitted these to the legislature … Most tax was levied once a year and there were, for the first time, records of how money was spent" (Downs & Larkey, 1986, p. 146). The work of the NYBMR on budgetary reform referred to above, paved the way for a different approach at federal level. President Taft was in favour of federal government adopting a similar budget structure, but the newly established Bureau of the Budget (BoB) located in Treasury could not garner sufficient traction when Taft was not re-elected (Downs & Larkey, 1986). Budgets remained structured along the lines of objects of expenditure (inputs) as opposed to activities of operating units (outputs) (Downs & Larkey, 1986).

Two interventions changed this. Firstly, the BoB was afforded greater power: its "day to day auditing and fiscal control tasks" were replaced by a managerial, more decentralised way of working made possible when the BoB relocated from the treasury department to the office of the Presidency in 1939 (Downs & Larkey, 1986:148). Secondly in 1949, President Truman, requested the Hoover Commission to create a "performance budget" (Cox, 2002, p. 164). Performance-based budgeting required budgets to be used not only as a mechanism to control spending, but also to consider allocation of resources based on the performance of programmes (McNab & Melese, 2003). The second Hoover Commission (1955) made recommendations on budgets, costs and management reports (Bouckaert, 1990).

A practical application of performance-based budgeting was found in the newly established RAND Corporation (1949). This organisation was created with the aim of producing an integrated air force

budget that pulled together all defence expenditure and civilian programme budgets under one umbrella (Downs & Larkey, 1986). Their recommendation culminated in programme budgeting, which entailed the development of a function-orientated budget that contained the expenditure of all relevant agencies and parties across four areas: strategic, tactical, defence and transportation (Downs & Larkey, 1986).

Not long thereafter, in 1965, President Johnson instructed all civilian agencies to implement the Planning, Programming, and Budgeting System (PPBS), with the first step being the development of cross-cutting, objective-orientated categories against which budgets can be appropriated (Downs & Larkey, 1986). The PPBS had a strong quantitative focus, as it also required each agency to cost their objectives which assisted policymakers in understanding how expenditure relates to objectives for the various agencies (Downs & Larkey, 1986).

Aside from the PPBS, several other performance-based budgeting reforms have been attempted, including: the Budget and Accounting Procedures Act, management by objectives (MBO) and zero-based budgeting (ZBB) (Jordan & Hackbart, 1999). A progression in performance-based budgeting approaches is evident when comparing the ZBB approach with its predecessor, the PPBS. The ZBB approach required departments to not only show how various levels of spending affect outputs (efficiency), but also how spending affects the measures of effectiveness (GOA, 1993).

None of these earlier performance-based budgeting reforms stood the test of time. The failure of these initiatives can be attributed to: i) conceptual and methodological difficulties in developing agreed-upon objectives, ii) difficulties in measuring these objectives, iii) the top-down imposition of these approaches leading to limited buy-in and commitment, iv) institutional shortcomings, such as limited human resources, insufficient agency capabilities, as well as outdated information systems to support the ever-growing data and reporting requirements (Downs & Larkey, 1996, Jordan & Hackbart, 1999, McNab & Melese, 2003, Posner & Fantone, 2007).

Aside from the budgetary related reforms geared at economy, productivity specific activities were also initiated, albeit with limited longevity and impact. The 1962 productivity project, led by the BoB, reviewed five government agencies' output and productivity, with the view of developing quantitative productivity measures (Hubbell & Kinghorn, 1988). Although their 1964 report showed that productivity measurement is in fact possible, and would provide the basis for rational decision-making, the BoB suggestions were never implemented (Downs & Larkey, 1986). The heightened interest in

budgetary reforms, diminished the interest in this project, leading to a re-assignment of staff to the PPBS (Downs & Larkey, 1986).

The productivity concept gained prominence again in the 1970s, this time driven by the political leadership. The motivation was not cost savings as per the previous time period, but rather ensuring "more yield out of the taxpayer's money" (Bouckaert, 1990, p.59). George Gilder, from the *National Commission on Productivity and Work Quality* equates the escalating attention to productivity with the expanding size of government – both in relation to its goods and service spend (estimated at 22% of GNP) as well as its growing employee base (Gilder, 1975).

US Senator Proxmire's concerns regarding the limited availability of efficiency measures at the federal level, resulted in the establishment of a productivity task force in 1970, consisting of representatives from the General Accounting Office (GAO), Office of Management and Budget (OMB) and Civil Service Commission (Hubbell & Kinghorn, 1988). The Joint task force started out by assessing the state of current productivity measurement, reviewing the reasons for its non-use. The task force developed productivity indices for selected activities, spanning 17 agencies, showing ironically that government was not nearly as unproductive as was previously assumed (Downs & Larkey, 1986).

Another federal level initiative was the establishment of the *National Commission on Productivity* in 1970 under the auspices of President Nixon. Although this commission financially supported the work of the Joint Task Force it was mainly concerned with private sector productivity (Bouckaert, 1990; Downs & Larkey, 1986). The commission had two name changes: In 1974 the commission became the *National Commission on Productivity and Work Quality* and in 1975 it became a recognised center of productivity tiled *the National Center for Productivity and Quality of Working Life* (Bouckaert, 1990). One of the center's legacies is the establishment of a dedicated journal on the topic of government productivity in 1975, called the *Public Productivity Review*. Unfortunately, after just eight years of existence, the center was closed in 1978 and its public sector functions were transferred to the *Center for Productive Public Management* (later called the National Center for Public Productivity) (Bouckaert, 1990).

Towards the end of the 1970s the productivity movement at a federal level started to slow down as other priorities moved to the fore. However, work on the municipal level and state level continued. Bouckaert (1990) attributes this continued interest to various government organisations (e.g. the National Center for Productivity), non-profit organisations (e.g. The Urban Institute), professional organisations (e.g.

the International City Management Association) and university research centres (e.g. State University of New York).

The 1980s saw an upsurge in the pursuit of productivity once again; this time sparked by the growing recognition of the value of private sector techniques in the public sector coupled with a demand for "less government" (Bouckaert, 1990:65). According to Bouckaert (1990), two trends in productivity emerged during this time: one more moderate approach that called for realistic expectations to be attached to productivity efforts and the recognition that productivity improvements by way of technological advancements and innovations can increase the costs of goods and services in the short term, but once implemented will lead to greater efficiency. The second stream of thinking was more radical: inspired by the private sector, government drastically needed to reduce the government deficits. The second approach took precedence and a "war on waste" was launched (Bouckaert, 1990:65). At a federal level, Reagan's private sector survey and the *Grace Commission* continued the emphasis on efficiency, with a strong focus on cost savings and reducing the growing deficit. This work paved the way for the NPM movement in the USA in the 1990s (Bouckaert, 1991, cited in Bouckaert & Halligan, 2006).

Both the GAO and the *President's Council on Management Improvement* advocated for an expansion of productivity efforts, recommending that the OMB lead productivity efforts at the federal level. The OMB agreed, setting both short- and longer-term goals around productivity (Bouckaert, 1990). Legislative backing for President Reagan's government-wide productivity improvement programme followed in February 1986, by means of the *Productivity Improvement Program for the Federal Government Executive Order 12552* (Hubbell & Kinghorn, 1988).

The emphasis on productivity was primarily concerned with the efficient operation of government employees but fell short in dealing with the consumer and its requirements in terms of effectiveness and the provision of quality services (Milakovich, 1992, p. 577). The Total Quality Management (TQM) movement addressed these shortcomings by taking a holistic approach to performance and paying more attention to customer satisfaction:

> Quality management is a strategic integrated management system for achieving customer satisfaction through the involvement of all employees and continuous improvement of all the organisations' processes and use of resources. All three stakeholders' agendas – that of customer, employee, funder, must be served with quality (US, FQI, cited in Holzer & Callahan, 1998, p. 35).

In the USA, TQM has its origins in statistical consultant W. Edwards Deming's application of statistical control theories in some of the largest companies. Other scholars recognised for their advancement of TQM include Walter Shewhart (1931), Armand Fiegenbaum (1954) and Joseph M. Juran (1954, 1979, 1988) (Milakovich, 1992). The TQM is viewed as an alternative to the closed rational performance measurement systems (i.e. MBO, PPBS), as it places quality at the forefront, thereby bringing a range of other organisational variables into play (Heinrich, 2003).

Outside the USA, TQM started gaining traction in the early 1950s in Japan when applying TQM as a means of improving the quality of their goods and services (Milakovich, 1990). The nine attributes of TQM, as set out by Burstein and Sedlak (1998, p. 124), are listed below:

- Top management is committed to quality and productivity;
- A customer orientation permeates the agency;
- Teamwork at all levels seen as key to improving service delivery;
- Quality management training provided at all levels;
- Accountability for quality and productivity improvement tied to performance evaluation;
- Recognition and incentive programmes established and used creatively;
- Measures and standards set for quality service delivery;
- Efforts underway to eliminate barriers to productivity and quality;
- Constant stimulation to improve quality and productivity.

At a local government level, quality was brought to the fore by the *Service Efforts and Accomplishment* initiative, run by the Governmental Accounting Standards Board (GASB). The GASB is an independent non-governmental organisation (NGO) that advises government on financial reporting and have been encouraging a focus on quality and outcomes (Wholey & Hatry, 1992). From a methodological perspective the emphasis on quality has led to the development of more sophisticated ways of measuring results, for example the development of satisfaction surveys to capture the multi-dimensional nature of quality (Bouckaert, 1990).

2.2.3 Contribution of PA to performance measurement

In summary, first generation productivity efforts commenced at the local government level with the NYBMR taking the lead. The USA local government history shows a clear evolution in performance measurement: initially the focus was on two of the 3Es (economy, efficiency and effectiveness): namely

economy and efficiency, residing under the banner of productivity (instead of performance). The underlying assumption was that if efficiency is "sorted", results would automatically follow and hence outcomes did not receive much attention initially. The scientific tradition significantly influenced the way measurement activities were undertaken: a scientific and value free approach was propagated, explaining the greater emphasis on the measurable aspects of economy and efficiency as opposed to effectiveness. Federal government level measurement activities were also concerned with efficiency and productivity but were influenced by the economic and political climate of the various time periods. This resulted in a "stop start" interest in efficiency and productivity initiatives, which impeded on the ability of these initiatives to produce long-lasting impact. In the 1980s, the quality dimension of performance started receiving attention, which led to the quality movement and the introduction of citizen satisfaction surveys within the public sector.

It is evident from the historical overview that performance measurement as a distinct concept does not feature during the PA. Instead, the various dimensions and goals of performance were being studied. This includes productivity, efficiency, economy and to a lesser extent effectiveness and quality. Efficiency took precedence, with the meaning attached to efficiency shaped by the non-partisan nature of government administration at the start of the twentieth century. Two divergent views around efficiency emerged: one whereby efficiency came to be "defined as the accomplishment of the work with the least expenditure of manpower and materials", and another less popular view that also considered the benefits that occurred (Bouckaert, 1992, p. 17). The former prevailed as this was aligned with the view that the administration needed to focus on its own "technicity" and the results would then follow.

Productivity was also an evolving concept: Initially Hatry and Fisk subscribed to a broader approach of productivity that encapsulates effectiveness *and* efficiency (Bouckaert, 1990). This understanding of productivity was however changed in 1977, when performance (instead of productivity) came to encapsulate efficiency *and* effectiveness. Productivity was subsequently more narrowly defined as the amount of output per unit of input, which represents an inverse relationship of efficiency, defined as cost per unit of output (Bouckaert, 1990).

The earliest reference to outcomes (which speaks to effectiveness) is found in the work of Walter F. Wilcox in 1896. He "advised that the benefit of a sewerage system should be measured in terms of decreased mortality rather than in terms of increased productivity" (1896, cited in Williams, 2003, p. 646). Williams (2004) viewed the period leading up to the 1920s as a stage where "the concern over results neither advanced nor retreated" (Williams, 2004, p. 151). Regardless, outcomes received some

attention as early as 1907 in William Allen's *Efficient Democracy* handbook, and other outcome related studies on education and health produced by scholars such as William Ogburn, Joseph Neff, Mabel Walker and Edison Cramer (Williams, 2003; 2004). Although an interest in measuring outcomes, the lack of effectiveness measures, resulted in workload measures being used to capture both agency productivity and outcomes of the agency.

The next historical section, the NPM, marks a pivotal point in the performance measurement literature. It is associated with a shifting role for government: requiring the official to move from administrator to assume greater managerial responsibilities. From a performance management perspective, the NPM ignited the performance measurement movement and facilitated a shift from outputs to outcomes (Pollitt, 2003).

2.3 The New Public Management (NPM)

The 1980s is associated with an upsurge in government reforms, specifically in the USA and the UK (Miller, 2005). The prevailing PA regime was being criticised for being impersonal and overly focused on "rigorous logic specification" (Runya et al, p. 11). The NPM movement diverted from the linear hierarchical bureaucratic structures, effecting both structural and procedural changes through the introduction of private sector concepts and market mechanisms into the USA and UK public sectors during the 1980s and 1990s (Ropret & Aristovnik, 2019, Runya et al., 2015).

Our attention will be focused on the significance that this movement offers in terms of performance measurement. I will start by providing a historical overview of some of the key performance related reforms in the UK and USA, followed by a broad overview of the core principles and drivers of the NPM.

2.3.1 The UK performance measurement related reforms: the financial management initiative

The turning point of the reform history in the UK is commonly associated with Margaret Thatcher's term of office, which started with the 1979 victory of the Conservative Party. Carter, Klein, and Day, (1992, p. 5) describe the Thatcher era as a "managerial revolution" which saw "a concerted effort to reform the public service along efficiency, effectiveness and market and private sector principles" (Miller, 2005, p. 15). Even though the emphasis is on the FMI, its "predecessors" – the Programme Analysis and Review (PAR) as well as the Efficiency Unit are covered briefly to set the scene in terms of the shortcomings these reforms attempted to address.

The first pre-FMI reform I will cover is the PAR. Through a joint arrangement (between Treasury, Central Policy Review Staff (CPRS) and departmental staff, PAR intended to bring policy closer to the financial considerations (Gray & Jenkins, 1982). The conduct of PAR did not go as planned. From a treasury perspective, the intent was to focus on programmes that had significant budgets assigned to them. However, departments viewed this as a threat to their programmes, and therefore, it became a challenge to find appropriate topics to review. The CPRS was interested in reviewing programmes and activities that cut across departments, but this also faced resistance as political interest and support was often lacking (Gray & Jenkins, 1982). Another criticism of PAR is the relevance and timing surrounding the delivery of reports to the Minister: "PARs seemed neither to fit ministerial timetables nor to correspond to ministerial interest. Thus, PAR became buried in a world dominated by the overall interests of the Whitehall system" (Gray & Jenkins, 1985, p. 113). By 1974 the programme started winding down and in 1979 it was formally discontinued (Gray & Jenkins, 1985).

The second reform, the establishment of an Efficiency Unit followed shortly after Thatcher came into office in 1979. A commitment was made to drastically reduce the size of government and managerial efficiency became the flavour of the day. Ms Thatcher looked to the private sector for solutions and subsequently appointed Sir Derek Rayner (former Chief Executive of Marks & Spencer) to head the Efficiency Unit (Carter et al., 1992). Despite their small size, the Efficiency Unit had direct access to the Prime Minister and was a key initiative in the accountable management era: "[T]he strategy has not been simply a low level, cost-cutting exercise. It is also associated with the beginning of broader changes, especially in administrative attitudes and accountable management" (Gray & Jenkins, 1986, p. 172). The efficiency reviews differed from the PAR, in that they were much more action oriented. Ministers received detailed steps of what needed to be done to reduce costs, thereby achieving greater efficiencies.

Several streams converged around the 1980s, making the time ripe for a more coordinated approach to performance measurement in the UK.

> There were, then, a number of different arguments pointing in the same direction – the need to develop measures of performance. On the one hand, there were the managerial arguments. There could be no way of assessing progress in achieving greater efficiency and effectiveness without such measures. Similarly, the decentralisation of government activities inevitably meant centralising knowledge about what they were doing. On the other hand, there were political arguments. If the constitutional fiction or ministerial accountability to Parliament was to be made a reality, then it was essential to devise

> currencies of evaluation that would allow MPs to scrutinise the departmental record. Lastly, there was an element hitherto no one touched on: accountability to the public (Carter et al., 1992, p. 19).

The FMI directly addressed the needs raised by Carter et al (1992). The FMI made two valuable contributions from a performance measurement perspective which will be further discussed below: firstly, it linked the origin of performance indicators to the FMI, and secondly, it heightened an interest in accountability which resulted in accountable management.

The term "performance indicators" is directly linked to the FMI manifesto with Prime Minister Thatcher proclaiming, "that a thousand PIs should flourish" (Carter et al., 1992, p. 2). The FMI required all government units to specify measurable objectives, as well as stipulate the required costs for each project or programme. Following this, performance indicators and output measures needed to be developed to track the achievement of objectives (Carter et al., 1992). However, these performance indicators did not focus on effectiveness, with outcomes not being mentioned and "the customer gets a passing glance" (Gray & Jenkins, 1986, p.185).

Carter et al. (1992) associate this tendency to focus on inputs with the complexities of government programmes. Departments with clearly defined functions were able to partake in the FMI much more extensively than departments with multiple or unclear objectives. Evidently, performance indicators during this period were often statistics disguised as performance indicators, most often reflecting the inputs of departments, bar the exception of some departments attempting to link outputs to targets, for example the Department of the Environment (Carter et al., 1992).

The second contribution of the FMI was the heightened interest in accountability, which in turn gave rise to the notion of accountable management, defined as follows by the Fulton Committee:

> Accountable management means holding individuals and units responsible for performance measured as objectively as possible. Its achievement depends on identifying or establishing accountable units within government departments – units where output can be measured as objectively as possible and where individuals can be held personally responsible for their performance (Fulton, 1968, p. 51, cited in Carter, Klein & Day, 1992, p. 11).

Although associated with Margaret Thatcher's term of office and the rise of NPM, the history of accountable management goes back to the Heath Administration and its attempts to rationalise government (Gray & Jenkins, 1985). A dual role was embedded in 'accountable management': not only is cost and budgetary information utilised to make decisions but also to give an account of performance

(Humphrey, Miller, & Scapens, 1993). Accountable management ascribes to a neoliberal agenda, which places a heavy emphasis on economic (financial) measures – a key characteristic of the reforms initiated during this era (Humphrey et al., 1993). This required government managers to take on an entirely new role as budget holder, requiring them to become financially savvy while simultaneously overseeing their operations (Gray & Jenkins, 1982).

The above also had implications for the way in which effectiveness and efficiency were applied within the context of FMI. Given the heavy emphasis on cost, efficiency came to be associated with reducing inputs at the expense of outputs. It is therefore not surprising that the heightened emphasis on financial management adversely affected the scrutiny of policy effects and strategic planning of the UK government. The era of accountable management is therefore input-dominated with a heavy emphasis on target-setting and holding managers accountable mainly for input targets (Gray & Jenkins, 1986).

I now turn to the history of performance measurement in the USA to cover two topics: performance monitoring and the focus on outcomes by way of the USA-NPM model. The Government Performance and Result Act (GPRA) marks the single biggest federal government level effort at instituting performance measurement (albeit under the guise of performance-based budgeting). This reform placed significant emphasis on outcome measurement through an introduction of non-financial indicators.

2.3.2 Performance related reforms in the USA: performance monitoring and the GPRA

The GPRA (1993) is associated with formalising the practice of performance measurement and evaluation in the USA Federal government, as well as advancing an outcomes-based approach to performance measurement.

Performance measurement in fact constitutes monitoring as can be seen from the GAO's definition: "Performance measurement is the systematic ongoing monitoring and reporting of program accomplishments, particularly progress towards pre-established goals or standards" (GAO, 2012 cited in Hatry, 2013, p. 23). The intent is for performance monitoring to extend beyond the measurement of inputs and outputs, to include program quality and outcomes (Wholey & Hatry, 1992).

The point must be made that "standard" Programme Evaluation practice constitute both monitoring and evaluation. However, from the above, it is evident that within the public sector context there is a clear separation of the monitoring and evaluation functions, with performance measurement encapsulating performance monitoring, and programme evaluation being solely associated with evaluation.

The outcome impetus to performance measurement (or performance monitoring) predates the GPRA. Our early history under the PA regime revealed that local government was the forerunner on measurement related efforts. This included the focus on outcomes with cities such as New York and Charlotte (North Carolina) undertaking the regular collection of outcomes-related data in the 1970s already (Hatry, 2013), as well as the work undertaken by the GASB, mentioned above.

Two other legislative efforts prior to the GPRA also warrant mention here: the CFO Act of 1990 and Senator William Roth's Bill (1991). Both these legislative documents addressed the development of performance measures. The CFO Act stipulates the development of performance measures, as well as requiring agencies to submit reports and financials that showed progress on results (Joyce, 1993). The Senator William Roth Bill titled "Federal Program Performance Standards and Goals Act of 1991" required federal agencies to submits plans around the development of performance indicators, that had to cover both outputs and outcomes (Wholey & Hatry, 1992). In addition, annual performance reports showing progress against these indicators had to be submitted (Wholey & Hatry, 1992).

The *Reinventing Government* publication by Osborne and Gaebler is directly linked to the introduction of the GPRA, calling on government to steer, not row. This resonated with President Bill Clinton and in 1993 he called for a review of government, named the National Performance Review, which was led by Vice President Al Gore (Joyce, 1993). The National Performance Review was inward focused and looked at the way government works, as opposed to the activities the government should be undertaking (Halachmi, 2002c). The review found: i) high occurrences of waste and inefficiencies in government and ii) inadequate attention to the formulation of programme goals and the associated measurement thereof. These findings gave effect to the GPRA, which was signed into law by President Clinton in 1993. The purposes of the GPRA were to improve the confidence of the USA people in government, improve federal programme effectiveness, improve service delivery, improve internal management and congressional decision-making as well as initiate programme performance reform (Halachmi, 2002c, p. 69).

The GPRA required government agencies to develop a series of documents: i) strategic plans, which should contain agency goals and objectives, ii) annual performance plans which operationalise the strategic goals in quantifiable terms through the use of performance indicators, as well as iii) annual performance reports that compared actual with planned performance (Joyce, 1993).

In summary, the USA-NPM model, which gained traction during Clinton's term of office, places the citizen centre stage. Influenced by Osborne and Gaebler's seminal work, *Reinventing Government (1992),* the focus turned to results, downsizing government and strengthening state and local level

agencies through decentralisation (De Vries & Nemec, 2013). The UK-NPM model swayed towards the business-type managerialism, borrowing principles from the private sector such as "hands-on management, performance measures, disaggregation and competition within the public sector and copying private sector management styles" (De Vries & Nemec, 2013, p. 6). Both the USA and UK NPM models entailed the adoption of performance management as a means of ensuring accountability for results. I elaborate on this and other core principles of NPM in the next section.

2.3.3 Drivers and core principles of NPM

As indicated above, the NPM movement "intended to enact a break from the traditional model of public administration underpinned by Weber's (1946) bureaucracy, Wilson's (1887) policy-administration divide, and Taylor's scientific management model (1911) of work organisation" (O'Flynn, 2007, p. 354).

Several drivers are associated with the introduction of NPM. In the previous section we already discussed the influence of the changing political landscape in the UK and USA and how the new conservative governments instituted market reforms: Margaret Thatcher through the Conservative Government in the UK in 1979, and the Republican Party's Ronald Reagan in the USA in 1978.

Other drivers associated with the introduction of NPM include:

- Citizens' waning trust in government and its ability to execute its functions efficiently and effectively. Greater accountability was viewed as a panacea to restore this trust;
- Economic and fiscal crises: this is one of the most commonly agreed instigators attached to NPM. Countries were increasingly faced with escalating costs due to the expanding size of government. This was attributed to government being the direct provider of a multitude of goods and services;
- The influence of neoliberal ideas: government, being the main provider of services, also resulted in it being viewed as a welfare state, but with increasing inefficiencies. The only way to rectify this situation was by means of market competition. Additionally, the lack of a proper reward system and dominance of self-interest in the public service, meant no incentives existed for government to control costs. The solution was either to introduce more rules or to create an additional structure to oversee government agencies. This did not solve the problem, and instead, resulted in an even bigger public sector;
- Technological advancement: the expanding innovations in information technology are viewed as an enabling factor to ensure accountability, despite a more decentralised approach. More

refined information systems had the advantage of keeping track of outside agency performance and hence strengthening accountability;

- The role of management consultants: international management consultants, accountancy firms and international financial institutions played a key role in importing private sector practices and bringing new management techniques into the public sector.

(Barberis, 1998; Hope, 2001; Larbi, 1999).

The key elements of NPM cannot be neatly contained and presented as a universally agreed list. Different scholars place the emphasis on different concepts or understandings of what NPM entails. The table below provides a summary of the main NPM doctrines, according to Hood (1995). He shows how the NPM doctrines have replaced the classic administration doctrines and the implication thereof on the accounting function (which includes performance measurement aspects).

Table 1: NPM doctrines and how it replaces traditional doctrines

NPM Doctrine	Replaces	Accounting implications
1. Unbundling of public services into corporatised units organised by product	Belief in uniform and inclusive public service to avoid underlaps[4] and overlaps in accountability	More cost centre units
2. More contract-based competitive provision, with internal markets and term contracts	Unspecified employment contracts, open-ended provision, linking of purchase, provision production to cut transaction costs	More stress on and identifying of costs and understanding cost structures; so cost data becomes commercially confidential and cooperative behaviour becomes costly
3. Emphasis on private-sector styles of management practices	Emphasis on public-sector ethic fixed pay and hiring rules, model employer orientation, centralised personnel structure, jobs for life	Private-sector accounting norms
4. More emphasis on discipline and frugality in resource use	Stable base budget and establishment of norms, minimum standards, union vetoes	More emphasis on the bottom line
5. More emphasis on visible hands-on top management	Paramount emphasis on policy skills and rules, not active management	Fewer general procedural constraints on handling of contracts, cash, and staff, coupled with more use of financial data for management accounting

[4] Underlaps in accountability occurs when some aspects of service delivery or performance measurements is not undertaken. Both the institutional design as well as lack of clarity around who is responsible are reasons why underlaps occur

NPM Doctrine	Replaces	Accounting implications
6. Explicit formal measurable standards and measures for performance and success	Qualitative and implicit norms and standards	Performance indicators and audit
7. Greater emphasis on output control	Emphasis on procedure and control	Move away from detailed accounting for particular activities towards broader cost centre accounting; may involve blurring of funds for pay and activity

(Hood, 1995 p.96)

Hood's sixth doctrine shows the move away from qualitative, implicit norms and standards to explicit, quantitative measures. Pollitt (2003, p. 27) in his summary of the key elements of the NPM, also highlights the shift from "inputs and processes towards outputs and results". The Organisation for Economic Co-operation and Development (OECD) (1995, cited in Hope, 2001) attributes the strong performance-orientated culture during NPM to the greater emphasis that was placed on results and the quality of service. Additionally, the increased decentralisation of services enabled service delivery to occur in closer proximity to the citizen, which in turn gave citizens an opportunity to provide feedback on the quality of services (OECD, 1995 cited in Hope, 2001). NPM also introduced principles such as marketisation and privatisation which ensured the most efficient, cost effective options were explored (OECD, 1995 cited in Hope, 2001). This emphasis on efficiency found expression through performance targets.

Criticism of NPM centre around the extent to which it can be viewed as a novel, fully-fledged movement as opposed to an incoherent set of ideas borrowed and rebranded from the private sector (Lane, 2000). At the heart of the criticism is the suitability of applying private-sector principles to the public sector (Radej, Golobic, & Cernic Istenic, 2010, Ropret & Aristovnik, 2019,). Cameron (2009) points to the faulty assumption that the private and public sectors share many similarities, hence the belief that the public sector should benefit from private sector approaches and practices. In line with this, Pollitt (1998, cited in Miller, 2005) argues that the narrow consideration of efficiency as the main concern negates other objectives that are important for governments to consider, for example, the concept of equity as well as the intended and unintended effects of policy outcomes.

2.3.4 NPM's contributions to performance measurement

The NPM regime saw the expansion of performance related concepts and terminology, such as performance indicators, performance measurement and performance monitoring. Performance measurement as an established practice in the UK is directly associated with the conservative

governments of the 1980s and 1990s (Boland and Fowler, 2000). In the USA, the GPRA (1993) formalised the practice of performance measurement in the Federal Government. In expanding the performance measurement terminology, the FMI initiative in the UK is associated with the term "performance indicators". The USA NPM model is acknowledged for advancing an outcomes-based approach. However, the emphasis on monitoring is evident with performance measurement being linked with performance monitoring, and evaluation as part of programme evaluation.

Performance, and no longer productivity became synonymous with the concepts of economy, efficiency and effectiveness (Talbot, 1999 cited in Bouckaert and Halligan, 2006). Overlaying these three concepts with the contemporary systems logic translates into the following:

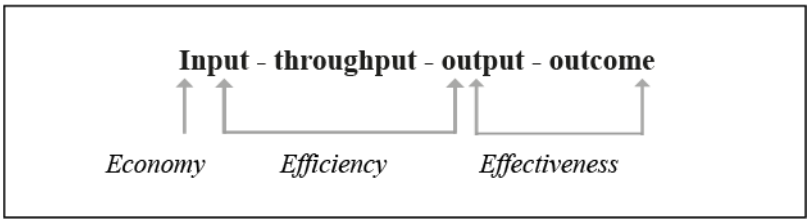


Figure 4: System model of organisations, with 3Es included

The recognised definitions for the 3Es are as follow (Boland & Fowler, 2000; Downs & Larkey, 1986):

- Economy is concerned with the costs of inputs such as staffing, project budget and operational costs. The principle of economy requires that inputs are minimised, for example switching to cheaper inputs;
- Efficiency is typically expressed as a ratio of input to outputs. The most used measures of efficiency are labour productivity and unit cost. An increase in outputs for a set level of inputs shows how efficient the organisations is in converting inputs into outputs;
- Effectiveness relates to the extent to which outputs meet the organisational objectives.

The following two definitions of performance measurement captures the key insights from the NPM regime:

> Performance measurement describes levels of performance in relation to some standard and is typically a univariate measure, whereas program evaluation enables the explanation of why certain levels of performance were observed and is thus multivariate, using a number of performance measures to support the explanation (McLaughlin & Jordan, 2010, p. 72).

> …[A]n expansive definition being performance measurement, the regular generation, collection, analysis, reporting and utilisation of a range of data related to the operation of public organisations and public programs, including data on inputs, outputs and outcomes (Bouckaert & Halligan, 2006, p. 5).

The definitions demonstrate the following:

- Performance measurement is associated with some standard or level of performance that needs to be achieved;
- Performance measurement is presumably less complex than programme evaluation, and can be satisfactorily measured by single measures (or indicators);
- Performance measurement is an ongoing activity, hence its association with performance monitoring;
- Performance measurement involves a series of steps from data generation to analysis and use of data;
- Performance measurement includes data collection on outcomes in addition to inputs and outputs.

2.4 NPG regime and performance measurement

Around the mid-1990s, European and American scholars signalled a change to the way government was running: Jan Kooiman at a 1993 symposium discussed the emergence of what he called 'Social political governance' to encapsulate the new interactions between government and society (Lynn, 2010, p. 110). Inspired by earlier works on governance, such as Osborne and Gaebler's *Reinventing Government* (1992), new governance arrangements were being proposed to address the issues of fragmentation and decentralisation in the public sector (Lynn, 2010). Additionally, this new way of doing business meant "citizens would be empowered rather than being merely the passive objects of government initiatives" (Lynn, 2010, p. 111).

The term "governance" dates back to the 14th century when it was used in France to mean "seat of government" (Loffler, 2003, p. 216). The term came to the fore again when used in a 1989 World Bank report to "signal a new approach to government" (Loffler, 2003, p. 216). Explaining governance has been undertaken in various ways: some authors approach governance as an umbrella term that captures different models and modes of governance (Hill & Hupe, 2014), while other scholars focus on a particular era within government, for example Toonen (1998, cited in Cloete, 2000), who links the

emergence of governance to the NPM era in government. (Toonen, 1998, p. 233, cited in Cloete, 2000, p.14).

One of the most common ways in which governance is defined is in relation to the steering role of government and utilising networks as a means of doing so: The definition provided by Bovaird and Loffler captures the plurality of actors involved in public sector service delivery: "We take public governance to mean "how an organisation works with its partners, stakeholders and networks to influence the outcomes of public policies" (Bovaird & Loffler, 2003, p. 6). Good governance with its normative connotation describes how governance should be undertaken. This definition from the United Nations provides quite an extensive list of requirements:

> Good governance has eight major characteristics. It is participatory, consensus oriented, accountable, transparent, responsive, effective and efficient, equitable and inclusive and follows the rule of law. It assures that corruption is minimised, the views of minorities are taken into account and that the voices of the most vulnerable in society are heard in decision making. It is also responsive to the present and future needs of society (Loffler, 2003, p. 217).

The elements commonly found in most governance definitions include the following (Loffler, 2003, pp. 217–218):

- The involvement of multiple stakeholders in problem solving, given the nature of societal problems being faced;
- Both formal (laws, regulations) and informal rules (codes of ethics, traditions) apply, but depending on the power dynamics, one can become more important than the other;
- Hierarchies and cooperative networks which replace the market structures of NPM as the key steering mechanisms;
- More focus on the key processes involved in social interaction, as opposed to the narrow concern with inputs and outputs.

As can be seen from Loffler's list, the rise of the NPG is directly associated with: i) the limitations government faces in resolving 'wicked' societal problems on their own, and ii) the need to address the shortcomings of the NPM, not just structurally, but also at a procedural level. The simplistic rationality that underpins private sector principles, and which have been seen in its introduction into the public sector by way of the NPM, were deemed to be unsuited for the public sector context (Radej et al., 2010).

NPM has left a legacy of fragmentation given the application of principles such as contractualisation, privatisation and marketisation (Christensen, 2012; Torfing & Triantafillou, 2013). The extensive intra-organisational focus of the NPM as a means of achieving greater efficiency was replaced by a new way of thinking in the late 1990s. To tackle the most complex societal problems, institutional boundaries must be transcended, and various networks activated to pool resources and efforts (Christensen, 2012).

Examples of the "wicked problems" faced by government include regional economic development, social inequality, climate change and the development of future-proof infrastructure (Koppenjan et al., 2019). Wicked problems are characterised by: i) a lack of a clear problem definition, ii) they are multi-causal, multi-scalar and interconnected, iii) multiple stakeholders with conflicting agendas are involved, iv) these problems tend to straddle organisational and disciplinary boundaries, v) solutions to wicked problems impacts on the broader system, vi) solutions are not necessarily right or wrong, simply better or worse, vii) it can take a long time to evaluate solutions, and viii) every problem is unique (Wahl, 2017).

The traditional approaches to dealing with uncertainties do not offer satisfactory solutions. For example, trying to simplify wicked problems by breaking them down into "more manageable elements" is an inappropriate means of dealing with these "untameable problems" (Head, 2018, p.7). Instead of trying to solve these problems, means should be found to manage these problems better. Collecting more information or initiating further research seldom solves divergent views or opinions on different subject matters (Koppenjan & Klijn, 2004). Similarly, applying top-down approaches in managing wicked problems only leads to an exclusion of key players, so creating resistance to the solutions being proposed (Koppenjan & Klijn, 2004).

Network and institutional theory offer a way of dealing with these uncertainties as it takes into consideration the double plurality of government: not only are multiple actors involved in public services (the notion of a plural state), but multiple processes contribute to the policymaking system (the notion of the pluralist state) (Osborne, 2010).

Two schools of thought can be distinguished in network theory (Koppenjan & Klijn, 2004). Firstly, there are the network theoreticians who focus on the institutional aspects, such as the process of network formation (for example Blau, 1982; Laumann & Knoke, 1987), network interdependencies (for example Aldrich, 1979; Rhodes, 1981) and closedness of networks (for example Rhodes, 1988). Secondly, there are those focusing on the strategic processes and management of networks. Networks are now viewed

as a pivotal element of the "institutional architecture" of many public sectors (Keast & Mandell, 2013, p.27).

The emphasis on greater coordination and a more holistic approach to government are doctrines associated with PA, albeit under a new guise of "joined up government" and "whole of government" (Christensen, 2012). In the UK, the notion of joined up government is associated with the modernisation programme (1999) and was viewed as a solution to addressing conflicting policy outcomes, ensuring better use of government resources, finding more innovative ways of working and providing citizens with more integrated access to government services (Hodges, 2012).

Aside from correcting the NPM shortcomings, the rise of NPG is also associated with the need to return power to the centre of government, which included restoring the political control that was depleted during the NPM (Christensen, 2012). This has given rise to many governments establishing inter-ministerial and inter-agency units as a means of coordinating cross-cutting priorities better (Christensen, 2012).

The overarching assumption of the NPG is that a collaborative working arrangement will lead to better performance. Other NPG assumptions, which could be viewed as risks to the "success" of this regime, include (Koppenjan & Koliba, 2013; Torfing & Triantafillou, 2013):

- The assumption that the various network actors have the desire to work together;
- That key stakeholders are able to work together, i.e. can set aside the time, have the means of engaging and participating in critical conversations;
- That government officials are willing, and able to take on this new role of network facilitator or network participator;
- The assumed presence of trust and collaboration among network stakeholders. Trust can enable greater flexibility around decision-making and implementation but should not be over-estimated in contributing to the achievement of policy objectives.

It is evident that NPG will require new skill sets, as well as new modes of problem identification and problem solving (Torfing & Triantafillou, 2013). Other risks associated with NPG are the high transaction costs associated with horizontal integration and coordination of stakeholders (Koppenjan et al., 2019).

2.5 The NPG's contribution to performance measurement

Complexity scholars question the suitability of the "results-oriented economistic solutions offered by NPM consultants" within the public sector (Kiel, 1994; Teisman et al., 2009, cited in Head, 2018, p. 11). Head (2018) says that the scientific and rational approaches to performance measurement as utilised during the previous regimes, notably NPM, are an inadequate response to these wicked problems. It is proposed that old forms of performance tracking be replaced with new forms of accountability that not only captures the collective goals and outcomes to be achieved, but also the procedural aspects of stakeholder involvement. Whereas NPM focused on the results of an organisation or unit, public governance "pays a lot of attention to how different organisations interact to achieve a higher level of desired results" (Bovaird & Loffler, 2003, p. 9).

This is aligned to the thinking of Voets, van Dooren and De Rynck (2008), who posit that performance measurement under the NPM were predominantly focused on the macro-level (country) and micro-level (organisation or individual), disregarding the intermediate network level where multiple actors interact. Our historic overview already alluded to the system approach, which Voets et al. (2008) refer to as a production logic whereby inputs are transformed into outputs and ultimately outcomes. This narrow view of performance with its predominant focus on the micro-level does not capture the multidimensional and multilevel nature of performance measurement under the "network society" (Voets et al, 2008, p. 774). An expanded performance measurement framework is therefore proposed by Voets et al. (2008) that includes the traditional NPM production performance, but with two added dimensions, namely process performance and regime performance.

For each of these dimensions, a list of criteria is provided, operationalising the three dimensions. Production performance entails the "traditional" measurement aspects of goal attainment and efficiency of the project or programme intervention (Voets et al., 2008). The process performance dimension offers means of measuring member legitimacy (i.e. members' commitment to network goals), accountability (which assesses the degree to which members are held accountable) and accordance (extent of network consultation and consent practices in the network) (Voets et al., 2008). Regime performance is measured through membership, the extent to which the network is institutionalised and the quality of the network relationships (Voets et al., 2008).

Another performance measurement model developed under the NPG is that of Keast and Mandell (2013). Their focus is on the interplay between the type of networks and the role of people in the network, leading to the development of a conceptual framework that draws on the commonly used

terminology within network theory: cooperation, coordination and collaboration. These three network types operate along a continuum and are expected to solicit different behavioural dynamics that affect the performance of the network. For example, at the lowest level of networking, i.e. cooperation, there is no agreement of a common goal, but rather a sharing of resources. Furthermore, network actors' behaviour would be characterised by infrequent communication with other actors; and accountability for performance residing within the respective organisations. Juxtaposed with this is the collaborative network structure where there is complete interdependence between the various actors in achieving the agreed-upon goal. A high degree of trust amongst the various actors are required to develop "new ways of thinking and behaving, … in their existing systems of operation and service delivery" (Keast & Mandell, 2013, p. 30). On the people side, the framework aims to capture the connection between the extent to which an individual identifies with the activities of the network, as well as their level of agency in enacting the work of the network, bearing in mind the network type.

Almquist (2012) in his summation of the implication of NPG for performance measurement, focuses on the broadened meaning of accountability compared to NPM. Accountability for results remain, but the means of deciding which indicators to track are done collaboratively with other network actors. In addition, indicators must also be developed that will assess the functionality of the network. Koppenjan and Koliba (2013) refer to this as the substantive issues (i.e. outcomes) and procedural issues (i.e. the functioning and quality of the network) of network performance.

The challenges associated with performance measurement can potentially be amplified within the NPG: not only do more actors need to reach agreement on the objectives to be achieved, a set of mutually agreed indicators also need to be developed that cover the performance of the joint project or intervention, as well as the performance of the network. Notwithstanding these concerns, it is evident that performance measurement continues to remain a central mechanism for measuring government performance under the NPG. This is confirmed by the growth in contract management that started under the NPM and continues under the NPG, in both the USA and the UK. Third sector organisations are increasingly being contracted to deliver critical government services (Smith & Smyth, 2010). To ensure accountability, performance contracting is instituted as a means of controlling the performance of the third sector organisations. A recent trend has been outcomes-based contracting, which requires service providers to demonstrate achievement of outcomes and impact, instead of only outputs (Smith & Smyth, 2010). These performance requirements place heavy demands on third sector organisations: not only are new systems and processes needed, the very purpose and culture of these organisations are often challenged, as explained:

> The shift to outcome evaluation often involves a revolution in thinking for agencies and their overall management. Agencies need new investment in management information systems and monitoring in order to track and compile important programmatic and financial data, track outcomes and better understand their cost structure. The effect is to professionalise the administrative infrastructure of third-sector organisations since a more substantial infrastructure can require new resources that may be at a variance with the previous programmatic focus of the organisation (Smith & Smyth, 2010, p. 278).

In closing, the three different public sector regimes, i.e. PA, NPM and NPG, should not be considered in isolation from each other. Even though these regimes support different doctrines, have different theoretical bases, and have advanced performance measurement in different ways, public sector reforms tend to be characterised by "combination, complexity, layering and hybridization, rather than by dominance, substitution and pendulum swings" (Christensen et al., 2007, cited in Christensen, 2012, p. 6).

Having concluded our history of performance measurement approaches in the public sector, I now discuss the challenges associated with performance measurement.

2.6 Challenges in performance measurement

Performance measurement places the work of government officials under the magnifying glass. It is therefore not surprising that a multitude of dysfunctional practices have emerged because of performance measurement. Smith (1995a) puts this down to four reasons (Table 2): i) misalignment between organisational objectives and indicators being tracked, ii) the challenges associated with measuring complex organisations, iii) the inability to use the data appropriately and iv) the lack of continual review to ensure continued relevance of measurement systems. Under each of these categories, different behaviours emerge. For example, tunnel vision is described as the situation where the organisation focuses on the things that can be measured, as opposed to the things that should be measured. Myopia occurs when there is a focus on short term targets as opposed to longer term objectives. Both these behaviours result in a misalignment between the performance measurement system and the organisational objectives. The complexity of societal problems that governments deal with also influence measurement behaviour. One example under this category of divergent behaviour is measure fixation whereby officials become so focused on the measure that the achievement of the objective or outcome becomes a secondary concern. Governments also struggle to process performance information correctly. One example of a divergent behaviour under this category is gaming, whereby

officials deliberately under-achieve so lower targets can be set going forward. The final category of divergent behaviours deals with the inflexibility of government performance measurement systems: ossification occurs when a performance measurement system has become outdated and non-responsive to the context. This not only leads to a manipulation of the system but limited value being added by the performance measurement system.

Table 2: Divergent behaviour and causal factors

Factors	Resultant type of behaviours
A divergence between organisational objectives and the measurement scheme	Tunnel vision (focusing on measurable, easy indicators)
	Sub-optimisation of individual departments or units to the detriment of the total system
	Myopia: focusing on short-term targets as opposed to longer term objectives
The challenges associated with measuring complex organisations	Measure fixation: focusing on the indicator rather than the desired outcome
	Misrepresentation: misreporting or distorting data to create a good impression
An inability to process performance data correctly	Misinterpretation as indicators are frequently statistical measures which means they are collated in a league table and there is no difference between them
	Gaming: deliberately under achieving to obtain a lower target next time
An inability to shift focus when context change	Ossification, so that when an indicator is no longer relevant, it is not revised or removed

(Smith, 1995a)

Perrin (2002) covers many of the same dysfunctional behaviours listed by Smith (1995a), but also refers to goal displacement. The following example illustrates the point: in order to reduce the number of patients on waiting lists for surgery, those who are the easiest to treat are assisted first in order to push up the numbers. Gaming, also listed above by Smith, is described as "hitting the target and missing the point" and can be broken down into three categories (Hood, 2006):

- Ratchet effects: deliberately setting low targets, knowing that the target in subsequent years will be based on current performance;
- Threshold effects: where target setting does not accommodate the better performing entities; in fact, those already achieving targets might decrease their outputs to meet the targets;
- Output distortion: where the achievement of targets become the "end all and be all" leading to data being adjusted, or even changed to ensure targets are met.

Different typologies have been developed to capture the multitude of challenges associated with performance measurement. Mayne (2007, p.90) distinguishes between three types of challenges: organisational, technical and conceptual:

> Organisational challenges cover areas where organisations and the people in them need to change or to do things not done before. Technical challenges are those where expertise is required in measurement and reporting. And for many of these challenges there are embedded conceptual challenges where new thinking is required about a problem.

Organisational support by means of top management buy-in is highlighted as the most critical success factor in implementing a performance measurement system (Berman & Wang, 2000; Cook et al., 1995; Mayne, 2007; Nielsen & Hunter, 2013; Theurer, 1998). If supported by the leadership, measurement efforts will receive the needed attention, have the required budgets assigned to it, and ensure that the data is utilised (Berman & Wang, 2000; Perrin, 2002).

In addition to leadership support, it is of critical importance that the performance measurement system aligns with the strategies and priorities of the organisation. This means that the indicators selected must measure the goals and outcomes of the organisation (Carter, 1990; Nielsen & Hunter, 2013; Theurer, 1998). Additionally, there needs to be clear ownership of the performance management system, ensuring that the performance data is used as a learning opportunity, instead of a punitive exercise to catch staff out (Nielsen & Hunter, 2013; Theurer, 1998).

Ensuring that performance measurement systems are aligned with the organisational strategies and objectives is easier said than done. Executing this at the organisational level is challenging for the following reasons: firstly, agreeing objectives across units and agencies is a mammoth task that ideally requires extensive engagement with multiple stakeholders. A distinction must also be made between the executive and administration who might want to pursue different objectives. Secondly, some objectives can be in conflict which each other, thirdly the meaning of objectives and outcomes are not well understood and finally, quantifying objectives and outcomes is extremely challenging and can result in the wrong things being measured. (Berman & Wang, 2000, Joyce, 1993)

Outcome measurement is challenging for several reasons. Firstly, it is difficult to quantify social phenomena. Contrary to the economic discipline, it cannot be reduced to one common currency or

denominator. Secondly, it is recognised that outcomes take time to materialise and often need to be undertaken or produced jointly across organisations (Carter et al., 1992; Smith, 1995b; Matheson, 2002). Thirdly, there are technical challenges relating to outcome measurement such as the pressure to demonstrate causal attribution between outputs and outcomes (Carter, Klein & Day, 1992; Mayne, 2007). Fourthly, outcomes are likely to invite public scrutiny which could be challenging to explain given the limited influence of government over the achievement of outcomes (Matheson, 2002). Finally, outcome measurement requires a fundamental shift within government, or as Behn (2002, p. 7, cited in Mayne, 2007) puts it: "[I]it requires a complete mental reorientation ... and it usually requires that elusive cultural change whereby performance information becomes valued as essential to good management".

Given these measurement challenges, it is important to have realistic expectations around performance measurement. Indicators can only convey part of the picture and caution must be applied in interpreting the data. Ridgway (1956) came to the same conclusion when discussing the dysfunctional consequences of performance measurement, proposing that multiple measures should be used. This, however, does not offer a full solution as a ranking will then be required to prioritise the indicators, and some of the indicators might contradict each other (Ridgway, 1956). Another limitation is that one is not able to gauge the overall effect given the lack of a single indicator. A composite score could help resolve this challenge, but significant thought will then need to go into a suitable weighting scheme to ensure the sub-goals are adequately reflected in the composite score (Ridgway, 1956). Solutions to address the technical challenges regarding indicator construction include: the development of standard definitions for all indicators (Cook et al, 1995), establishing whether an indicator could be susceptible to manipulation, whether the indicator is unambiguous and easily understandable and making explicit the indicators limitations (Carter, 1991). Finally, it is imperative that indicators are reviewed continuously to ensure continued relevance and improvement (Cook et al., 1995; Perrin, 2002).

Given the complexity in measuring outcomes, one runs the risk of developing fragmented performance indicator systems that are "short term, unmanageably large, and may reflect the preoccupations of only some of the stakeholders" (Smith, 1995b, p. 16). This statement could leave the reader feeling pessimistic about the value of performance measurement in determining government effectiveness.

However, I will argue (following some other recent authors for e.g. Nielsen & Hunter, 2013) that programme evaluation as a distinct field of enquiry, has much to offer as a complementary approach to performance measurement to overcome some of the outcome measurement challenges discussed above.

Programme evaluation complements performance measurement (or performance monitoring) in multiple ways. At a high level, programme evaluation offers a deeper consideration of the why and how questions. Through case studies and comparative analysis, the evaluation profession can shed some light on the more complex questions around government performance (Newcomer (1997). Nielsen and Hunter (2013), drawing on the classifications of Rist (2006), and Nielsen and Ejler (2008), identify five ways in which programme evaluation can enhance performance measurement:

- Sequential complementarity: because monitoring usually precedes evaluation, monitoring processes can identify key questions to be answered by evaluation studies;
- Information complementarity: both M&E activities require data; if aligned, then the same data sources and systems can be drawn upon to address both M&E;
- Organisational complementarity: rather than having M&E dealt with by two units, organisations can ensure both functions are dealt with by one unit;
- Methodological complementarity: very similar approaches, tools, processes and systems are utilised to generate data for M&E purposes;
- Hierarchical complementarity: performance information within a government system can be utilised at different levels, for example national data can be used as a benchmark for local governments.

Blalock (1999) expands on the methodological complementarity raised by Nielsen and Hunter (2013): during the conceptualisation phase evaluation professionals utilises tools such as the theory of change and logic models to clarify goals and outcomes. And in the measurement stage, instead of selecting the easily measurable indicators, relevant indicators are formulated, and assistance provided with the development of suitable performance standards and targets.

Many countries are now progressing towards the combined usage of programme evaluation as part of the broader set of performance information. Examples include the USA, Canada and Australia. The Obama Administration, through the OMB and the first Federal Chief Performance Officer encouraged agencies to conduct programme evaluations (McDavid et al., 2013). In Canada, policy prescribes the undertaking of evaluations of all programmes in a five-year cycle (McDavid, et al., 2013). Recent reform in the Australian public sector has also committed to utilising both performance evaluation and performance measurement as a way of assessing government performance (Guthrie & English, 1997).

Learning from best practice abroad, South Africa has followed a similar path. Performance measurement and programme evaluation play complementary roles in ensuring that a comprehensive picture of government performance is obtained. The next section expands on the development of performance measurement in South Africa.

2.7 Performance measurement in South Africa

Following the abolishment of Apartheid in 1994, the African National Congress (ANC)-led government undertook a major overhaul of the public sector as "prior to 1994 much of the population was excluded from service delivery, services to citizens were fragmented by ethnicity and there was no integrated system for data or measuring performance" (CLEAR, 2012, p. 145). The reform agenda needed to enhance accountability, while addressing the needs of the citizens: the "authoritarian, repressive and oligarchic" state had to be replaced with one that is "democratic, developmental and committed to a culture of human rights" (Cameron & Tapscott, 2000, p. 81). The newly elected government developed the *White Paper on the Transformation of the Public Sector* (WPTS, 1995), listing several imperatives for the new public service:

- Restructured and rationalised in order to become unified and integrated, but also more cost effective;
- Management practices and organisational culture had to be addressed in order to achieve the institution building and management goals;
- The public sector had to be representative of the population of the country, leading to the implementation of affirmative action;
- Ensuring professionalism and accountability was enacted through the establishment of various government bodies such as the Public Protector, Auditor General and the Public Sector Commission. It was also prescribed that Director Generals will be held accountable via performance measures. Professionalism was advanced through the introduction of a code of conduct for the public service;
- Service delivery had to be transformed. This required national and provincial departments to undertake significant planning around their service delivery mandates, in terms of what it sets out to achieve, how it will be tracked and resourced. TQM and customer satisfaction became critical elements. The *Batho Pele* white paper also strengthened the notion of the public as "customers" and involving communities in service delivery. (Summarised from Miller, 2005, pp. 71–79)

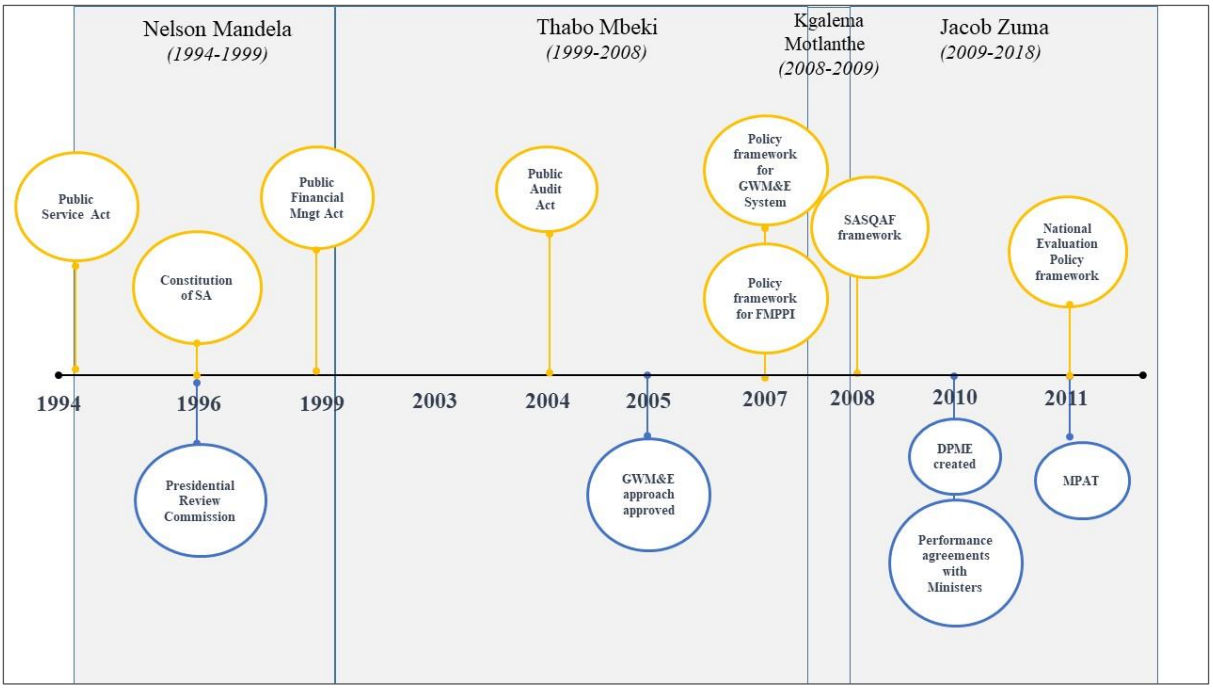
The prevailing public administration ideology, NPM, influenced the thinking around public sector reform. According to Mimba et al. (2007), the uptake of NPM within developing countries was driven by the need to demonstrate sound governance, as this is a prerequisite for organisational performance. Hope (2001) adds to this by identifying four other drivers of NPM within the developing country context: firstly, the economic or fiscal crises that have plagued many African countries since the mid-1970s and government's role in the economy and the direct provision of services; secondly, political instability that detracted from government's ability to operate in an accountable and transparent manner; thirdly, the use of complex institutional mechanisms and bureaucratic procedures that prevented the effective implementation of policy; and fourthly, the influence from international reform initiatives undertaken during the 1980s.

Even though Cameron (2009) questions the extent to which NPM was ultimately implemented, it is evident that NPM principles made its way into the South African public sector.

> [It] is generally accepted that NPM reforms were influential in South Africa. Miller (2005:70) states that much of the reforms (in South Africa) paralleled those which were implemented in other countries, in particular Britain and the USA. The Director-General for Public Services & Administration, Richard Levin (2004, p. 12–13) argues that public sector reform in South Africa has been shaped by the tenets of NPM, including a strong focus on decentralised management of human resources and finance (Cameron, 2009, p. 914–915).

One of the key tenets of NPM, as discussed earlier, is performance measurement. Even though other reforms have followed the NPM, performance measurement has not only "outlasted New Public Management, but also appears to escalate once in place" (Pollitt et al., 2010, cited in Jakobsen et al., 2017 p. 1). This is also the case in South Africa.

Figure 5 provides a historic timeline of the key initiatives, events and policies enacted to institute performance measurement in the South African public sector. It is by no means a comprehensive account of all the events and policy documents drafted in support of a more efficient, effective and accountable government. Yet, its purpose is to provide the reader with a "headline" view of how performance measurement evolved in the South African public sector since Apartheid. For the period under review, four different presidents headed the country since 1994: Mr Nelson Mandela, Dr Thabo Mbeki and Mr Jacob Zuma. Mr Kgalema Motlanthe served as President of South Africa for less than a year from September 2008 to May 2009, following the resignation of Thabo Mbeki.



Key: Yellow = policies and legislation, Blue = initiatives or key events

Figure 5: Major performance measurement policies, legislation and initiatives under the different presidents

A brief description of the performance measurement-related aspects of the policies and legislation included in Figure 5 are set out in the table below.

Table 3: Performance measurement related policies and legislation

Policies and legislation related to performance management	Purpose of policies or legislation
1994 Public Service Act and regulations	• The Public Service Act (1994) saw the establishment of three spheres of government. The Act also addresses staff appointments and managing staff performance. • The 1999 set of regulations introduces performance agreements for senior officials and sets a framework for managing performance of heads of departments (HoDs). • The 2001 regulations established a clear link between the staff development and departmental strategic goals as a means of driving greater organisational efficiency and accountability for results.
1996 Constitution of South Africa	• Separates government into three spheres: local, provincial and national, as well as stipulates the powers and functions of these spheres. • Sections 92, 133 and 195 of the Constitution address issues around government performance by way of the 3Es (efficient, economic and effective use of resources), encouraging greater transparency by making information available to the public and putting structures and stipulations in place surrounding lines of accountability.
1999 Public Finance Management Act (PFMA)	• Regulates financial management in national and provincial government. This included non–financial performance with accounting officers needing to report against predetermined objectives. • Paved the way for the development of strategic plans and annual performance plans (started in 2000 for provinces and 2010 for the national government). Also ensured a shift from inputs to outputs.
2004 Public Audit Act, Act 25 of 2004	• Legislates the auditing of performance information for all three spheres of government. The Auditor General is the responsible body. • An annual audit report is produced that assesses the quality of performance information, the accompanying evidence as well as the quality of performance information systems.
2005 GWM&E framework	Government approves a plan to implement a "system of systems" that prioritised a functional monitoring system, bearing in mind the existing M&E capacities.
2007 GWM&E Policy Framework	• This policy document expanded on the development of the systems included in the GWM&E: frameworks for programme performance information (FMPPI), quality of statistical data and evaluations. • Following from this, policy documents were produced for the three sub systems: o FMPPI (2007); o Statistics SA: South African Statistical Quality Assessment *Framework (*SASQAF) (2008); o The Presidency: National Evaluation Policy Framework (2011)

(CLEAR, 2012, Koma & Tshiyoyo, 2015, Minnaar, 2006, Phillips et al., 2014; Roos, 2012)

The presidential review commission (PRC) was established in 1996 to review existing transformation efforts, as well as make recommendations for the way forward. Their report, titled *Developing a culture of good governance*, was presented to President Nelson Mandela in 1998, which described the state of

the public sector as follow: "slow pace of transformation, the lack of a clear vision for change, the lack of effective leadership, ineffective strategic management and a lack of alignment between planning and budgeting" (Cameron & Tapscott, 2000, p. 82, Minnaar, 2006). Additionally, intergovernmental relations were found to be ineffective, despite the establishment of structures such as the inter-governmental forum (IGF) and Ministers and Members of Executive Councils Meetings (MINMECs) (Cameron & Tapscott, 2000). The report proposed an "integrated performance approach" for the public sector (Minnaar, 2006, p. 181).

The GWM&E system encapsulates all aspects of performance measurement within the South African context, giving effect to the integrated performance approach as identified by the PRC in 1996. Cloete (2009, p. 298) identifies several push factors for the development of a GWM&E system.

- A need to report back on the UN Millennium Development Goals[5];
- A lack of a national M&E system even when South Africa was hosting the World summit on sustainable development in 2002;
- No platform to provide feedback to citizens about the programme of action of government;
- Increased pressure from donors for more systematic assessment of programmes;
- The importance attached to M&E systems worldwide in enhancing governance.

In 2005, Cabinet approved the development of a GWM&E system. A draft policy framework for the GWM&E system was subsequently released in 2007, which described the three data terrains of this all-encompassing performance measurement system. This included: i) programme performance information, ii) statistical data and iii) evaluation. Following from this, three separate policy documents have been developed setting out the detail of each data terrain (DPME, 2011; Mouton, 2012):

- FMPPI in 2007 which describes the alignment of performance information from all three spheres of government, the role of performance information in planning, budgeting and reporting, guidelines in constructing performance indicators and clarification of key concepts;
- SASQAF in 2008 with the first edition of this framework providing the dimensions against which data quality and statistical products are assessed;
- National evaluation policy framework in 2011 which sets out to institute "a minimum system of evaluation across government" with the aim of promoting quality evaluation and ensuring results are used to improve government performance.

[5] Now the Sustainable Development goals

The need to bring about greater coherence between the three agencies in charge of these sub-systems was also highlighted (i.e. the presidency, national treasury and the national statistics agency) in the 2007 policy document. Various other stakeholders were also integral to the execution of the GWM&E system – for example the Auditor General's office, which annually audits all performance information, as well as the Department of Cooperative Governance (now called Department of Cooperative Governance and Traditional Affairs) who at that stage was responsible for local government performance monitoring. Figure 6 provides a summary overview of the critical M&E stakeholders, on the basis of their constitutional, legal and executive powers (Goldman et al., 2012, p. 2).

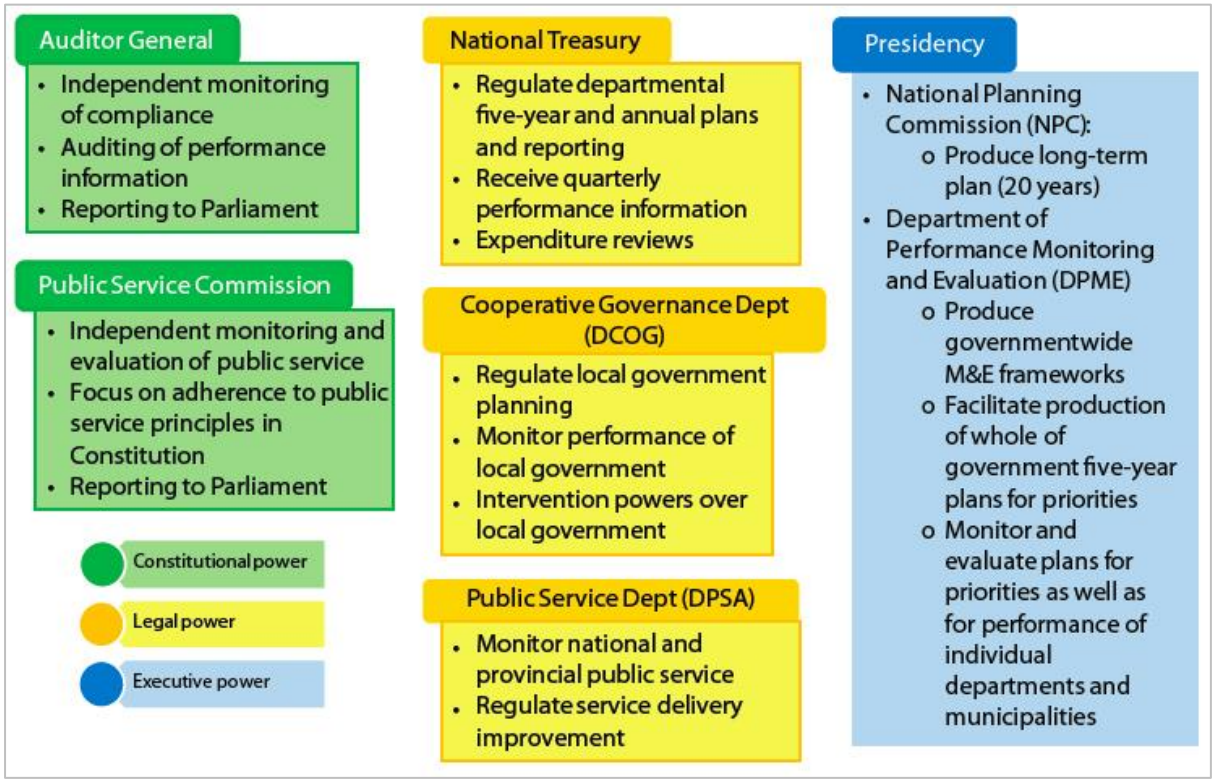


Figure 6: M&E stakeholders in South Africa

In 2009, the ANC was re-elected to office, but the significant loss of voter support, persistent problems around poverty, inequality and poor service delivery paved the way for a greater focus on M&E (Goldman et al., 2012). This led to the establishment of a Ministry of Performance Monitoring and Evaluation (MPME) within the Office of the Presidency, headed by Collins Chabane (Friedman, 2011). Minister Chabane tasked Ketso Gordhan (former Johannesburg city manager) and Ronette Engela (a monitoring and evaluation specialist) to provide solutions to the prevailing lack of government performance (Friedman, 2011). The team's starting point was the identification of key challenges impeding on government performance. These challenges were found to be three-fold: a) a lack of

accountability, b) poor planning and c) ineffective coordination between ministries especially when cross-cutting policy outcomes needed to be implemented (Friedman, 2011). Various countries were visited, and a desktop review undertaken to extract best practices. The UK Prime Minister Delivery Unit (PMDU) attracted much interest, with both Goldman et al. (2012) and Phillips et al. (2014) citing the influence of the UK-based PMDU in drafting the 2009 outcomes position paper titled "*Improving Government Performance: Our Approach*". The Department of Performance Monitoring and Evaluation (DPME) was established in January 2010 to ensure the execution of the outcomes based approach.

The outcomes based approach bears significant resemblance to the practices and principles of Deliverology. For example, Deliverology places a heavy emphasis on the selection of a few strategic priorities, as well as instituting routines by which to track these priorities. In line with this, 12 cross-cutting national outcomes were developed that fully aligned to the manifesto of the ANC (Phillips et al., 2014) and coordinating structures such as implementation forums introduced to support the achievements of outcomes (Goldman et al., 2012, p. 3). These forums, (or clusters as they are referred to now) bring together various departments and spheres of government to review quarterly progress reports (Goldman et al., 2012; Mabelebele, 2006; Phillips et al., 2014).

Similar to the UK PMDU, performance agreements were signed with all ministers by April 2010, holding them accountable to measurable targets (Friedman, 2011, Goldman et al., 2012)). Deliverology also prescribes the development of detailed delivery plans, against which performance can be tracked. The same was done locally - detailed outcomes based delivery plans were developed through a joint effort between representatives from government and civil society (Friedman, 2011). The intent with the delivery plans was to enhance performance measurement as demonstrated by this quote from Gordhan: 'We are going to tie you down to a number, to something that can be measured. If there was an innovation, that was it." (Friedman, 2011, p. 5)

In tracking these cross-cutting outcomes, the DPME had to take on several different roles (some evolving over time), which included assessing the management performance of departments, developing a national evaluation policy, implementing robust M&E related to the achievement of outcomes, undertaking hands-on monitoring and identifying appropriate support strategies for poor performance of local government (Umlaw & Chitepo, 2015, p. 2). Although good in principle, there are many challenges surrounding this collaborative way of working. These include power dynamics between different departments, a lack of data systems to produce the required data, a continued focus

on process as opposed to outcomes, inadequate use of data to improve performance and challenges translating delivery agreements into departmental plans (Goldman et al., 2012). Mabelebele (2006) concurs with this, stating that although coordination is taking place at a strategic level within these implementation forums or clusters, practicalities such as budgetary allocations and inter-departmental coordination still impedes on implementation.

Two more recent initiatives that exhibit characteristics of the Deliverology approach also warrant mention here. The first is Operation Phakisa (which means "hurry up" in Sesotho) (DPME, N.d.a). Launched in 2014, this programme draws on the "Big Fast Results" methodology applied by the consulting arm of the Malaysian Delivery Unit, PEMANDU (World Bank, 2017). This methodology was deemed suitable for accelerating some of the National Development Plan's priorities, starting with marine transport and manufacturing, offshore oil and gas exploration, aquaculture as well as marine protection services and ocean governance. The "Big Fast Results" methodology consist of eight steps and culminates in detailed implementation plans with ambitious targets and clear lines of responsibility (DPME, N.d.a). The plans were developed collaboratively with stakeholders from private sector, public sector and civil society (DPME, N.d.a). These working sessions are referred to as laboratories with dedicated time set aside to conclude the planning phase (six weeks in the case of Operation Phakisa). Once implementation starts rigorous monitoring is done, and any blockages resolved timeously.

The second initiative is the establishment of the Gauteng Delivery Support Unit in 2016. The Gauteng Delivery Support Unit is located in the Planning division of the Gauteng Office of the Premier and set out to institute the Deliverology approach in all of Gauteng's provincial departments with the distinction made between main priorities and secondary priorities (Gauteng Provincial Treasury, 2019). Main priorities were identified in the departments of education, economic development, community safety, human settlements, health and public transport, whilst secondary priorities were identified in the remaining eight provincial departments (Gauteng Provincial Treasury, 2019). Each provincial department was required to develop goal statements as well as the strategies that would ensure the achievement of these goal statements by 2019 (Gauteng Provincial Treasury, 2019). Following on from this service delivery agreements were subsequently signed between the Premier and relevant Members of the Executive Council (MEC) (Gauteng Provincial Treasury, 2019). These delivery agreements include specific targets and timeframes to enhance accountability and accelerate delivery.

It is evident from the above that many efforts have been undertaken to institute a performance measurement culture in South Africa, mainly under the auspices of the GWM&E system and through initiatives such as Operation Phakisa. In addition to standardising the practice of evaluation these efforts

and initiatives also ensures a greater emphasis on delivery and implementation –the focus of our next chapter.

2.8 Summary

Our overview of performance measurements in the USA, UK and South Africa provides valuable insights along four themes: i) the origin of performance measurement activities, ii) the rationale for performance measurement, iii) the widening scope of performance measurement over time as well as iv) the persistent challenges facing performance and outcome measurement.

Origin of performance measurement

In the USA, performance measurement started at the local level at the beginning of the Twentieth Century through the work of the NYMBR. Although federal level activities commenced soon thereafter, its intensity and impact were less than that found at the local level. This is attributed to the fact that federal level activities occurred quite haphazardly and were driven by different agendas during the PA regime. It was only under the NPM in the UK and USA that performance measurement finally settled at the federal level. In South Africa, performance measurement was driven at the national government level through the introduction of the GWM&E system. The system addresses all spheres of government, but with a clear champion in place through the DPME; its adoption in South Africa has been relatively quick and seamless.

Rationale for performance measurement

At a broad level, performance measurement aims to improve the performance of government. Digging deeper, however, we see how different political narratives have shaped the rationale for performance measurement. Under the PA in the USA, three sub phases can be distinguished: at the start of the century, the measurement of government was motivated by the need for better government. Corruption at local government level was rife and needed to be addressed. A clear separation of power between the administration and politicians were enacted as a means of achieving a more efficient government. The second phase under the PA covered post World War II (1940s to 1970s), when budgetary reform took precedence following a severely constrained economic environment. Budgetary initiatives such as the PPBS, MBO and ZBB were initiated during this time. In the 1970s, a third phase of greater political involvement in productivity ensued. With the interest of the citizen in mind, various efforts were undertaken to get productivity and efficiency back onto the agenda.

With the onset of the NPM in the UK and USA in the 1980s, significant advancements in performance measurement occurred. The political commitment from Thatcher and Clinton led to the widespread

adoption of performance measurement within government. In the UK, this was motivated by the dire economic environment, with emphasis placed on the introduction of performance measures through the FMI. In the USA, the shortcomings of financial indicators were increasingly recognised, which led to a greater focus on results as introduced by the GPRA. Under the NPG, the scientific and rational approaches in government practice are challenged; positing that the wicked and complex societal problems can only be meaningfully tackled when government and outside stakeholders work together. Performance measurement needed to take cognisance of the procedural aspects of the network, as well as capture the results of this collective effort.

In South Africa, performance measurement followed a rapid trajectory. Exposure to international best practice and a strong political drive to address the inequalities of the past, accelerated the introduction of performance measurement locally.

Widening scope of performance measurement

The three public sector regimes show a clear progression in the understanding and the scope of performance measurement. Under the PA, attention was afforded to the individual performance goals of efficiency, economy, productivity, and to a lesser extent effectiveness and quality. The influence of the scientific tradition is evident throughout the PA regime, given the dominant interest in the tangible aspects of government operations. Initially the concept of productivity included both efficiency and effectiveness, but in the 1970s, performance became the umbrella term for the 3Es of economy, efficiency and effectiveness, and productivity took on a narrower definition.

The NPM regime broadened the scope of performance measurement considerably. The GPRA not only formalised the practice of performance measurement and evaluation in the USA, but also advanced an outcome based approach to performance measurement. Contemporary performance measurement terminology is also associated with the NPM. The term performance indicators was first used as part of the FMI (UK based), while performance measurement in fact constitutes performance monitoring, which should not only include inputs and outputs but also short and medium term outcomes.

Under the NPG, performance measurement is further expanded to not only focus on the results (or outcomes) of the network, but also the functioning of the network. Two frameworks, covering different aspects of network performance have been discussed. Even though the NPG challenges the limitations of the rational and scientific era in government associated with the rise of performance measurement,

it still utilises the practice of performance measurement. The addition now is a) the number of stakeholders involved, and b) the added criteria of assessing the performance of the network.

Prevailing challenges of performance measurement and outcome measurement

The difficulties associated with performance measurement is a common thread that runs through the discussion in this chapter. Since the start of the previous century, attempts have been made to measure outcomes, and to link this to government spending. Outcome measurement only gained widespread traction through the enactment of the GPRA in the USA in the 1990s. Despite the spate of reforms, and commitment to results, governments continued to focus mainly on output measurement (and by implication, efficiency). The reasons for this relate to the conceptual, technical, and methodological challenges associated with outcome measurement (e.g. formulating clear objectives and outcomes, lack of capacity and systems to support outcome measurement, as well as the scientific expertise required to employ sound data collection methods and methodologies).

Performance measurement has prevailed through various public sector regimes. Initially, a small-scale endeavour, with a narrow focus, it has grown to become widely practised in countries across the world. Although some challenges remain, governments continue to use performance measurement as a means of demonstrating accountability and tracking progress.

With this historical overview of performance measurement concluded, I now move to a discussion on policy implementation, the other narrative with direct relevance to Deliverology.

Chapter 3: Policy implementation and the rise of implementation research

3.1 Introduction

Poor policy implementation and state capability are closely related with the World Bank defining state capability as "the ability to undertake and promote collective actions efficiently" (World Bank, 1997, p. 3, cited in Cloete, 2000, p. 10). Andrews, Pritchett and Woolcock (2016), using a combination of measures (averages of World Governance Index, Quality of Government and Failed State Index[6]), assessed 102 historically developing countries on their capability to implement policy. Their study demonstrated the "big stuck in state capability" with more than a third of all national studies showing low and deteriorating state capability. The authors argue that significant effort is geared towards policy design, but that implementation lacks the necessary attention: "… implementation failures hold back many countries from realizing their own stated development goals and that even worse, many governments lack the capability to overcome repeated implementation failures even after years of reforms designed to strengthen state capability" (Andrews et al., 2016, p. 7).

South Africa, although in the middle range of state capability (i.e. neither weak nor strong), is facing serious deterioration which cannot be ignored (Andrews et al., 2016). The emphasis on new policy and organisational structures, as opposed to what the policies and structures are supposed to do, is increasingly being raised as one of the key constraints towards improved government performance. Coupled with this, is the acknowledgement that most problems faced by governments are "wicked-hard – they are simultaneously logistically complex, politically contentious, have no known solution prior to starting, and contain numerous opportunities for personal discretion" (Andrews, Pritchett & Woolcock, 2015).

Effective implementation is a core element of a capable state.

[6] World Government index consists of six components. The authors extracted averages for "government effectiveness", "control of corruption" and "rule of law".
Quality of Government Institute provides a measure derived from the International Country Risk Guide data that is the simple average of three indicators: corruption, law and order and bureaucratic quality.
Failed State Index measures countries on 11 indicators related to likelihood of conflict. Only one indicator "Public services" were extracted. This indicator measures functions such as policing and criminality, infrastructure, roads, water, etc.

Various disciplines have studied the nature and processes of implementation, these include policy implementation research (as a distinct subject in the political sciences and public administration literature), programme evaluation and implementation sciences.

Programme evaluation considers implementation from two perspectives: firstly, programme monitoring, which is a systematic and continuous activity to ensure implementation happens as planned, but also on a periodic basis when process or implementation evaluations are undertaken. Process or implementation evaluations aim to determine whether the programme is reaching the targeted population and whether the project roll out is consistent with the programme design (i.e. the plan) (Rossi, Lipsey & Freeman, 2004). A breakdown in implementation is referred to as 'implementation failure' and will impact on the achievement of both outputs and outcomes. Rossi et al. (2004) distinguish between three types of implementation failure: "no intervention, or not enough, is delivered; second, the wrong intervention is delivered; and third, the intervention is unstandardized or uncontrolled and varies excessively across the target population" (Rossi et al., 2004, p. 191). Dimensions when studying implementation in this context include a consideration of, for example, fidelity, dosage, quality of programme delivery, participant responsiveness and programme reach.

Another implementation-related field that has attracted considerable interest during the past few decades is implementation science. In fact, a dedicated open access journal on Implementation Sciences was launched in 2005. Implementation science is an outflow of the evidence-based movement in medicine and the natural sciences and aims to "use research findings more effectively in routine clinical practice to develop a more research-informed practice" (Nilsen, Stahl, Roback & Cairney, 2013, p. 1). One of the key differences between policy implementation and implementation science is their scope: policy implementation covers a vast array of phenomena ranging from the implementation of quite simple and concrete policies to more complex and long-term policy imperatives. Implementation processes typically include many actors, cover different levels of government and can take years to effect change. Implementation science, on the other hand, focuses on clinical practices which have well-defined strategies, which, when implemented, can render outcomes within a short time frame (Nilsen et al., 2013).

This chapter will focus on policy implementation, given that the selected case (the eLearning Game Changer), is located within the public sector.

As with Chapter 2, I utilise Osborne's three public sector regimes as the framework. Showing the full history of policy implementation is important for two reasons: a) I will demonstrate how the three generations of implementation research link to the NPG as well as b) show how the NPG approach to implementation differed from the NPM; and in fact came about as a response to the shortcomings of the NPM, which include shortcomings related to implementation. Figure 7 summarises the overlap between the policy implementation research scholars and the three public sector regimes.

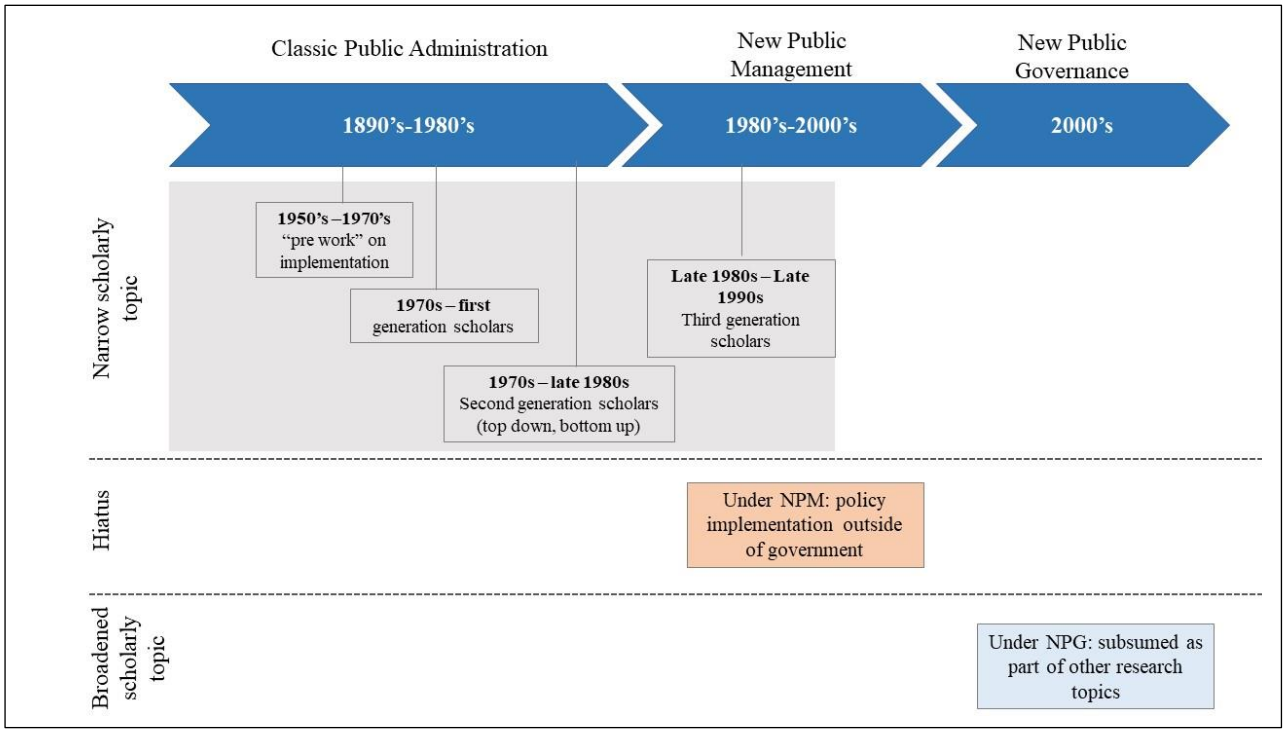


Figure 7: Alignment between public sector regimes and policy implementation

It is important to interpret the time periods as a broad indication of start and end dates and not absolutes. The dates are based on publication dates by leading scholars associated with the three generations of policy implementation research, recognising that other publications have been potentially produced earlier or after these specific dates. It should therefore not be viewed as the exact end or start of the three generations of policy implementation work but rather as a guideline to demonstrate the alignment to the three public sector regimes.

Figure 7 also makes a distinction between policy implementation as a narrow scholarly topic (grey shaded area above dotted line), the hiatus in policy implementation during the NPM (the orange shaded block) and when it became a broadened scholarly topic (the blue shaded block).

With the three public regimes as the framework for this chapter I will start with the time period when policy implementation was primarily studied under the narrow banner of theory development (mainly during the PA).

3.2 Implementation under the PA

The PA regime started in the late 19th century and extended through to the late 1970s/ early 1980s (Osborne, 2010). This era in government has several distinct characteristics: firstly, there is a clear divide between politics and administration which translated into the clear separation between the policy formation and policy implementation phases (Torfing & Triantafillou, 2013). Secondly, it was premised that government functioned within a closed, rational system whereby implementation followed on from the development of policy objectives in a linear manner (Osborne, 2010). This era is therefore characterised by a compliance to standardised procedures, a separation between policy formulation and implementation and a political and administrative dichotomy (Runya et al., 2015, p. 12).

It is during the PA, that policy implementation became a scholarly interest as will be seen through the three generations of scholarly work.

3.2.1 An introduction to policy implementation research

> Ours is an era of considerable pessimism and concern about government … Today's pessimism and concern stems from government's seeming inability to deliver, even in the case of programs with strong public backing which have been legitimately enacted into law … the concern is not with the objectives of the programs per se; rather, even if the programs are enacted, the ability to translate stated goals into reality may be beyond the capacity of government as we know it (Mazmanian & Sabatier, 1989, preface).

In line with Mazmanian and Sabatier's recognition that governments battle to implement policy, Palumbo and Calista (1990, pp. xi–xii) say that implementation, as a distinct step in the public policy process, has been ignored and undervalued in relation to the achievement of outcomes: "Over time, implementation researchers found that policy outcomes were not only shaped by the implementation process itself, but, in some instances, were actually determined by it".

Implementation research is predominantly concerned with what happens after a policy has been formulated with the view of developing a universal theory of successful implementation. It is for this reason that implementation research is mainly undertaken under the heading of implementation theory (Schofield, 2001). Implementation research started out as a predominantly North American endeavour, with Western Europeans joining the debates in the 1980s (Sabatier, 1986). The emergence of this sub-discipline is commonly associated with the failure of the Great Society social policies of the 1960s (Goggin, Bowman, & Lester, 1990; Van Meter & Van Horn, 1975), and the seminal work of Pressman and Wildavsky, titled *Implementation (1973)* (Hill & Hupe, 2014; Goggin et al, 1990; Mazmanian & Sabatier, 1989; Nilsen et al., 2013). This publication highlighted "the complexity of joint action" in implementing an economic development programme in Oakland, California (Winter, 2003).

Mazmanian and Sabatier (1989) also link the greater interest in implementation analysis to other changes in the public administration landscape during the 1950s and 1960s. Public administration studies during this time started questioning the commonly held belief that administrators seamlessly carried out legislative decisions, pointing to the external factors and multiple pressures facing the administration (Mazmanian & Sabatier, 1989). In addition, the development of systems theory sparked interest not only in the inputs of government, but also the results of government activities (Mazmanian & Sabatier, 1989). Saetren's research also found scholarly roots of implementation dating back to the 1950s: "[B]y the time *Implementation* was published in 1973, close to thirty books and over two hundred journal articles and doctoral dissertations respectively that employ implementation or implementing as a title word had been published or defended!" (Saetren, 2005, p. 569).

Many debates surround the field of implementation research, which include

- A definition of implementation that accurately reflects the wide scope and multiple actors involved in implementation;
- Agreement on what constitutes policy success or failure;
- The locus of control or authority over implementation, and with that, the extent of autonomy of other actors in the chain. Palumbo and Calisto (1990, p. 7) argue that organisations are social enterprises and that "administrators, no less than legislators, seek power, esteem, and monetary rewards that invariably shape how (or which) goals are achieved". Implementation is by its very nature political, as it does not occur in the absence of relationships and dependencies amongst actors;
- The appropriateness of drawing a clear distinction between the policy design and policy implementation phase which has its origins in the traditional Wilsonian perspective, and which draws a clear distinction between politics and administration (Palumbo & Calista, 1990).

The result of these theoretical and methodological debates has been the development of numerous frameworks and approaches to understand and explain policy implementation (Nilsen et al., 2013). Three generations of implementation research are discussed below.

3.2.2 First generation implementation research

The first generation research was undertaken before and during the 1970s and was mainly qualitative and a-theoretical in nature (Winter, 2003). Although the first generation scholars attempted to develop a theory around implementation, they are criticised for being overly pessimistic by focussing largely on policy failure when constructing implementation theories. For the first generation scholars "implementation success or failure is a function of flawed or imperfect primary legislation and a failure of bureaucratic compliance" (Schofield, 2001, p. 249). These scholars employed a largely rational, linear approach to theory building, focusing on very few variables (the number of actors and decision points), as well as the validity of the causal theory (Winter, 2003).

The works of Pressman and Wildavsky (1973), Bardach (1977) and Hargrove (1975) are placed within the first generation implementation research category.

- The commonly cited 1973 *Implementation* publication painted a bleak picture of the effectiveness of a federal economic development programme undertaken in Oakland as part of the "Great Expectation" movement after World War II (Hill & Hupe, 2014, p. 47). Pressman and Wildavsky (1973) focused mainly on the various role players involved in the implementation of the programme which is also reflected in their definition of implementation: "[T]he ability to forge subsequent links in the causal chain so as to obtain the desired results" (Hupe, 2011). Pressman and Wildavsky continues to draw a direct link between the number of players in the vertical chain and the ability to reach the required end result: the more players in the vertical chain, the smaller the chance of success, given the multiple opportunities for decisions to be vetoed and for a break in the chain to occur. They also introduced the notion of an "implementation deficit" which arises when a breakdown occurs between stakeholders or players in the vertical chain of delivery (Hupe, 2011).

- Eugene Bardach in his 1977 book, *Implementation Games* focuses on the conflict that occurs after policy adoption, positing that the political games continue even outside the public sector as implementing actors pursue their own interest (Winter, 2003). Bardach proposes pre-empting the variety of games played when programmes are being structured and also proposes a full "follow through" to mitigate the effect of these games (Hill & Hupe, 2014).

- Erwin Hargrove, in studying the policy process, recognised the limited attention afforded to implementation. He coined implementation as "the missing link" (1975), calling for more research on this area to fill this gap (Palumbo & Calista, 1990).

The second generation research is characterised by the development of analytical-theoretical frameworks and can be divided into two schools of thought: the top-down and bottom-up approaches to policy implementation (Hupe & Saetren, 2015).

3.2.3 Second generation implementation research: top-down approach

The majority of first generation research, as well as the initial work of the second generation implementation researchers subscribed to a top-down approach (Sabatier, 1986), and started in the early 1980s (Winter, 2003). The second generation phase marks a shift from explorative studies to theory development by constructing frameworks that empirically test which variables ultimately impact on successful implementation (Schofield, 2001).

The first characteristic of the top-down approach was that the policy decision or policy goal constituted the starting point. Matland (1995) refers to this as the "authoritative decision", with implementation then constituting the process of achieving a prescribed end goal. Berman, as well as Mazmanian and Sabatier's definitions, capture the authoritative nature of the top-down approach as well as policy success being associated with the achievement of policy impact. "Implementation analysis is, in short, the study of why authoritative decisions (policies, plans, laws and the like) do not lead to expected results (Berman, 1978, p. 160)."

> To understand what actually happens after a program is enacted or formulated is the subject of policy implementation: those events and activities that occur after the issuing of authoritative public policy directives, which include both the effort to administer and the substantive impacts on people and events (Mazmanian & Sabatier: 1986, p. 4).

The principles of "rule of law" and democracy are strongly embedded within the top-down approach, O'Toole (2004, p. 314) says that "top down analysts often express themselves clearly in support of a representative regime and the consistent execution of choices made by political leaders. They view any other position as a hijacking of democratic principles".

A second characteristic of the top-down approach was the clear distinction between the policy design (formation) phase and implementation phase. This view has its origin in the traditional Wilson politics-administration dichotomy introduced by Woodrow Wilson in the mid-1880s as referenced in the introductory section. In essence, this doctrine supports a clear separation between the politicians and administrators. For the top-down theorists, this was a critical distinction. Sabatier (1986, p. 31) motivates the need for this distinction on the basis of ensuring there are distinct points that can be evaluated as this allows to attribute success or failure to either the politicians or civil servants.

Some of the scholars singled out by Hill and Hupe (2014) in advancing the top-down approach, and their methodological contributions are as follows: Donald van Meter and Carl van Horn are credited for their system building approach whereby implementation success is mainly dependent on the degree of change required and the level of consensus that exists around the stated goals. Their model subsequently includes six clusters of variables that impact on the performance of a policy (Van Meter & Van Horn, 1975). These six variables are: i) policy standards and objectives (with the former needing to be concrete and specific), ii) resources and incentives available for implementation, iii) quality of the inter-organisational relationships, iv) characteristics of the implementation agencies, v) the economic, social and political environment and vi) the "disposition" or "response" of implementers which includes their understanding of the policy, the response to the policy (positive, negative, accepting), and the intensity of the response (Van Meter & Van Horn, 1975).

Hogwood and Gunn's contribution lie in the development of a list of recommendations for policymakers when setting policies (Hill & Hupe, 2014). Their recommendations, which can be viewed as an utopian situation, stipulate the need for adequate resources to be made available in support of the proposed programmes, the need for a sound causal theory to be present already during the policymaking stages, limiting the number of implementing agencies (ideally only one) to ensure simplicity, ensuring buy-in to the set goals, having the ability to "enforce" compliance of the set objectives, as well as having good communication in place amongst all parties (Hill & Hupe, 2014).

Mazmanian and Sabatier developed a framework that consists of three sub-sets of variables that impact on implementation: factors affecting the tractability of the problem, non-statutory variables affecting implementation and the ability of the statute to structure implementation (Mazmanian and Sabatier, 1989). The Mazmanian and Sabatier model is discussed in more detail below as an example of the top-down approach.

3.2.4 Mazmanian and Sabatier's model as an example of the top-down approach

Mazmanian and Sabatier's contribution is cited extensively in the top-down implementation literature. Winter (2003, p. 2013) refers to their work as the "best known and most frequently" used framework, while Matland (1995, p. 146) refers to this as the "most fully developed top down model". It is also noteworthy that the authors, Mazmanian and Sabatier, and others have tested their model in various settings to provide a critical self-reflection of its elements (Sabatier, 1986). The Mazmanian and Sabatier framework (Figure 8) consists of three categories of variables that impact on the implementation process. Across the three categories a total of 16 variables are discerned. The diagram below captures the framework, with notes provided below to explain the 16 variables (Mazmanian & Sabatier, 1989, p. 22).

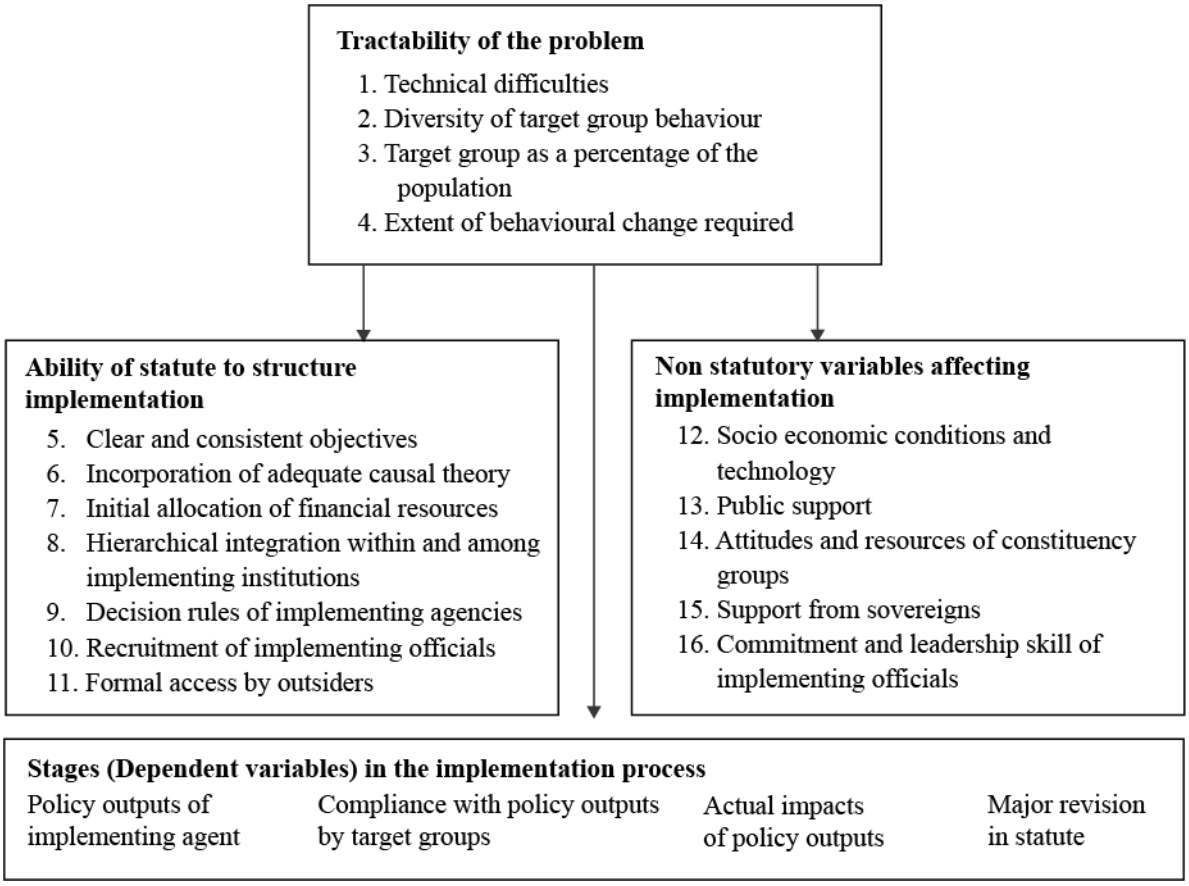


Figure 8: Mazmanian and Sabatier top-down framework

Notes explaining the variables (Mazmanian & Sabatier, 1989):

1. Technical difficulties: the development of adequate, cost effective performance indicators is an important prerequisite to the achievement of a goal as it provides clarity as to whether one is on track or not. Another technical challenge is the availability of technology as some interventions require a specific technology to be available at a cost-effective price.
2. Diversity of the target group behaviour: not only is it challenging to draw up succinct legislation and policies for certain behavioural aspects, it is almost guaranteed that implementation of these policies will result in variable implementation (and hence performance).
3. Size of the target group: affecting change within a small well-defined group will be easier than for a bigger undefined group.
4. Extent of behaviour change required: the more fundamental or far-reaching the behavioural change required, the more challenges one is likely to encounter.
5. Clear and consistent objectives: clarity of objectives helps focus resources and efforts.

6. Adequate causal theory: successful implementation is heavily dependent on a causal theory that clarifies the causal linkages. Secondly, it is critical that implementers have influence (jurisdiction) over a number of the critical linkages.
7. Initial funding: Funding is needed to implement a programme. At the start additional funding might be required to hire staff or to conduct preliminary work.
8. Hierarchical integration within and among implementing institutions: getting different spheres of government to work together is extremely challenging. Both sanctions and incentives can be utilised to ensure commitment and buy-in of all parties.
9. Decision rules: further rules can be introduced in implementing agencies to supplement the use of sanctions and incentives.
10. Officials' commitment: the success of any programme is heavily dependent on the commitment of officials. Unfortunately, it is not always possible to be prescriptive of which staff to employ.
11. Formal access by outsiders: not all target and beneficiary groups have the same level of legal standing: regulatory programmes invariably make it easier for citizens to raise their unhappiness in a court environment. In other arenas, such as consumer or environmental protection, beneficiaries as individuals will find it much more difficult to petition against adverse decisions.
12. Socio-economic conditions and technology: any change in the socio-economic and technology landscape can easily shift the focus to other social problems.
13. Public support: public support tends to be cyclical as interest in the problem comes and goes. Public opinion is utilised often to steer the political agenda and to gain support for particular policy positions.
14. Attitudes and resources of constituency groups: keeping the attention and financial support from other constituency groups is a key variable of implementation.
15. Support from sovereigns: this includes the organisations in control of the legal and financial resources of a particular policy priority. Keeping these sovereign organisations on board is one of the biggest challenges, but so is navigating potentially conflicting policy goals and guarding against waning support.
16. Commitment and leadership skills of implementing officials: highlighted as the variable that has the most direct impact on successful implementation, this speaks to implementing officials' commitment to the goals and their ability to implement the policy.

In terms of the stages, the first stage is the production of outputs which are aligned with the policy objectives. The second phase highlighted by Mazmanian and Sabatier (1989), captures the extent to which the target group complies with the policy outputs. A variety of factors influences their compliance: amongst other the likelihood that non-compliance will be detected and the severity of the repercussions in the case of non-compliance. The next phase moves into the impact stage and looks at

whether the desired outcomes were achieved. It is during this phase that the unintended outcomes often surface. The final stage is the revision or reformulation of a statute to reflect the extent to which impact has been obtained, to reflect any recent policy changes, as well as consider recent socio-political changes. Even though presented as a static framework, the Mazmanian and Sabatier model is in fact a dynamic model with multiple linkages between and across the 16 variables.

In summary, Mazmanian and Sabatier provide their "checklist" of success factors.

- The enabling legislation or other legal directive mandates policy objectives which are clear and consistent or at least provides substantive criteria for resolving goal conflicts;
- The enabling legislation incorporates a sound theory identifying the principal factors and causal linkages affecting policy objectives and gives implementing officials sufficient jurisdiction over target groups, and other points of leverage to attain, at least potentially, the desired goals;
- The enabling legislation structures the implementation process so as to maximise the probability that implementing officials and target groups will perform as desired;
- The leaders of the implementing agency possess substantial managerial and political skill and are committed to statutory goals;
- The programme is actively supported by organised constituency groups and by a few key legislatures (or a chief executive) throughout the implementation process, with the courts being neutral or supportive;
- The relative priority of statutory objectives is not undermined over time by the emergence of conflicting public policies or by changes in relevant socio-economic conditions which weaken the causal theory or political support of the statute.

(Mazmanian & Sabatier, 1989, p. 41)

The point around sound causal theory warrants further elaboration. Palumbo and Calista (1990), in identifying some of the reasons why an implementation gap occurs, highlights the fact that in many instances legislation passed can be viewed to be symbolic, i.e. there is a willingness to respond to a problem, but not necessarily to solve the problem (Winter, 1990). Winter therefore includes the policy formation process in his model as a variable, positing that often policies are introduced that are "impossible to implement from the outset" and therefore doomed for failure (Winter, 1990, p. 23). Ensuring the developing of sound theories and hypotheses is one way to address this, as it forces consideration of the causes and consequences of society's wicked problems (Winter, 1990).

The influence of the PA on the first and second generation scholars are evident: first generation scholars premised that implementation failure was a direct result of either flawed policy or non-adherence to standards (Schofield, 2001) which can be linked to the scientific paradigm described in Chapter 2. Second generation top down scholars supported the clear divide between policy formation and policy implementation, a direct result of the prevailing administration/political divide as introduced by Wilson in the late 19th century. In addition, the development of systems theory within government during the PA sparked interest from implementation scholars not only in the inputs of government, but also the link that exists between implementation and results (Schofield, 2001).

Not all contexts lend themselves to a top-down approach. Some of the criticisms include questioning the centrality of government as the key decision-maker, neglecting the role of the private sector, street level bureaucrats and other local implementing agencies. Similarly, the level of consensus present during the policy formation is an important predictor for future roll out. There is also the recognition that where no dominant policy or lead agency existed, the top-down approach becomes challenging to implement, given the lack of a coherent causal theory and limited hierarchical integration. The street level bureaucrats' contribution and role are also underestimated in the top-down approach. These front-line staff members' "know how" and knowledge of the context offers a wealth of knowledge and should be tapped into. The clear distinction drawn between the policy formation and policy implementation phases by the top down theorists are forced and negates the reality that this is a fluid process with no clear start and end point. (Matland, 1995; Palumbo & Calista, 1990; Sabatier, 1986; Winter, 2003).

There is also disagreement on goal achievement being the ultimate dependent variable, given the limited influence over outcomes and the challenges surrounding operationalisation of policy goals. This stem from the fact that the policy design phase does not necessarily produce clear, unambiguous goals and objectives, leading to challenges during the implementation stage. (Matland, 1995; Palumbo & Calista, 1990; Sabatier, 1986; Winter, 2003)

The bottom-up theorists address some of these concerns by providing an alternative perspective to implementation.

3.2.5 Second generation implementation research: bottom-up approaches

Bottom-up theorists, many who originate from Europe, placed emphasis on the local context by shifting "the analytical attention away from variables at the top or center of the system to the contextual and field variables at the bottom" (Nilsen et al, 2013, p. 2). For them, the evaluative standard is the problem

statement (e.g. unemployment), as opposed to the policy goal as is the case with the "top downers" (Nilsen et al., 2013; O'Toole, 2004). They also offer a wider definition of implementation by taking into consideration the work done outside of government agencies and building upon these existing strategies (Sabatier, 1986). Their emphasis on networks of actors has been a major "methodological contribution to implementation analysis" (Schofield, 2001, p. 251). The role of the street level bureaucrat is recognised in the policy-formation stage with these theorists viewing upfront policy formulation to be a small part of the process, instead supporting the view that "much policy is made during implementation itself" (Palumbo & Calista, 1990, p. 11).

A number of scholars' contributions to the bottom-up approach warrants mention here. Berman (1978) draws a clear distinction between macro- and micro-implementation problems facing implementers, positing that ultimately most of the power resides at the local level. Macro-implementation happens at a federal level and spans across a multitude of mandates, actors, agencies and role-players. All of these organisations or role-players have their own goals, resource challenges, communication methods and spheres of influence. Ultimately, a policy passes through this "macro structure" in a "step-like fashion" (Berman, 1978, p. 167), starting with the administration stage whereby a policy gets translated into a government programme, followed by the adoption of a local project as part of the adoption stage. Next comes the micro-implementation phase which is where mutual adaptation according to Berman (1978, p. 172–173) will determine whether the final stage of outcome achievement (called technical validity) will happen.

Michael Lipsky popularised the idea of the street-level bureaucrats and the vital role they play in policy implementation. According to Lipsky, these local bureaucrats apply their discretionary powers to adjust the policy to fit the citizen. In the process however, the bureaucrat is faced with multiple demands and high workloads which can distort the original intent of the legislation, in particular where more skills and efforts is required (Winter, 2003).

Hjern, in conjunction with others, made various contributions. Hjern with Hull emphasised the role of local networks and the identification of all contributing actors in addressing the problem statement (Sabatier, 1986). Through their snowball method and a sociometric method, Hull and Hjern mapped the "informal, empirical implementation structure around a given problem" (Winter, 2003, p. 214). These structures stretch beyond organisational boundaries and therefore tend to be less hierarchical and more informal. Hjern, in collaboration with Porter, looked at policy implementation from the perspective of organisational theory, contributing new thinking to the way organisational structures are perceived in the implementation process (Schofield, 2001).

Barrett and Fudge, as British scholars, built on Hjern and his associates' work by critically reflecting on the way organisations work. Aside from hierarchical orders, Barrett and Fudge also raise strategies such as compromise, negotiation, manipulation, and coercion as a means of getting the job done (Hill & Hupe, 2014). Another contribution of Barrett and Fudge has been a revised definition of implementation to ensure it reflects a focus on performance as opposed to conformance (to goals and objectives).

Elmore, in attempting to better understand the peripheral contextual factors, conducted a backward mapping exercise. The starting point constitutes the impact or goal to be achieved, mapping backwards both the implementation processes, as well as the actors involved in executing these processes (Schofield, 2001). Not surprisingly, this process helped to identify the multitude of actors involved in implementation.

However, the bottom-up approach has also attracted criticism. First, the potential exists for substantial deviation from the original policy goals as ultimately only the elected representatives, through a democratic process, have the power to set policy. Transferring too much power to the street-level bureaucrat could result in a sub-standard level of implementation which impedes on the ability to meet citizens' needs. Secondly, too much autonomy at the local level also affects implementation as ultimately the more active, well-resourced actors would be in the driver seat. On the positive side, Sabatier (1986), in reflecting on Hjern's work, highlights several positive points of the bottom up theorists. Bottom uppers have the advantage of addressing a societal problem not only from the perspective of government, but also involving other actors and tackling the problem on many fronts. Linked to this is the ability, given the collaborative approach, to identify other results or outcomes, including unintended consequences. (Sabatier, 1986).

The bottom up second generation scholars' work does not show as clear alignment with the principles of the PA when compared to the top down second generation scholars. For example, the bottom up scholars challenge the strict divide between policy implementation and policy formation, as well as the fact that the policy objective constitute the ultimately authoritative decision. In addition, the role of the centre of government is of less relevance compared to the importance of the street level bureaucrat and other actors. This results in the introduction of both field variables and contextual variables in the bottom up theorists' work.

The legacy resulting from the second generation era of work is mainly inductive in nature with three common categories of predictor variables emerging in the literature: policy form and content, capacity

of the organisation(s) and their resources, as well as the qualifications of people implementing projects (Goggin et al., 1990). Despite this categorisation of variables, O'Toole, in reviewing approximately 100 implementation studies, finds that the majority of studies "merely identify important variables presented as a checklist or in a theoretically oriented discussion of implementation process, without specifying a fully-developed model of implementation (O'Toole, 1986, p. 184). The problem of "too few cases, too many variables" (Lijphart, 1971, cited in Goggin, 1986) is one of the reasons for the emergence of the third generation research. Goggin (1986, p. 342) proposes a focus on the following: "identifying key variables, selecting cases more purposefully, combining small and large N studies, and conducting experiments and quasi-experiments". Other shortfalls in implementation research up to this point include the lack of quantitative measures associated with the many variables identified in the respective frameworks and models. Additionally, the bulk of second generation research focused only on federal and local level implementation, ignoring the role of the state (Goggin et al., 1990).

Recognising the shortcomings around theory building and empirical research on implementation, Winter (2012, cited in Hupe & Saetren, 2015, p. 95) proposes a different approach to implementation studies:

> Implementation research can be improved by 1) accepting theoretical diversity rather than looking for one common theoretical framework and 2) developing and testing partial theories and hypotheses rather than trying to reach for utopia in constructing a general implementation theory.

With debates between top-down and bottom-up scholars subsiding, implementation analysis was ready for a new generation of work. This call for greater synthesis, and a more scientific approach to implementation analysis, marks the start of the third generation paradigm.

3.2.6 Third generation research

The third generation research paradigm emerged during the latter half of the 1980s (Nilsen et al., 2013) and encapsulates a number of defining features in an attempt to advance implementation theory building in a more scientific manner (Goggin, 1986; Goggin et al., 1990; Lester et al., 1987, cited in Saetren, 2014):

- Key variables must be clearly defined;
- Hypotheses derived from theoretical constructs should guide empirical analysis;
- More use of statistical analysis using quantitative data to supplement qualitative analysis;
- More comparison across different units of analysis within and across policy sectors;
- More longitudinal research design.

Various frameworks were developed during this time to pull together the contributions from the previous two schools of thought (bottom uppers and top downers). Some of the combined models referenced in the literature include Winter's Integrated Implementation model, Goggin et al.'s Communication Model and Matland's Ambiguity-Conflict model. The latter two will be discussed below.

Matland (1995, p160) developed an ambiguity or conflict four quadrant model which offers four implementation perspectives (Figure 9). The level of conflict is directly related to the extent of the interdependence of various stakeholders, as well as the extent to which agreement exists on the objectives to be achieved (Matland, 1995). Ambiguity on the other hand steps in when there is a lack of goal clarity and when the means of achieving the goals are not clear (for example, when the respective stakeholders are not clear on their particular roles) (Matland, 1995). His four quadrant model results in various types of implementation processes as well as the key principle driving the achievement of outcomes (Matland, 1995). For example, in a low conflict, low ambiguity context the availability of resources will be the key determinant of successful implementation. However, in the case of high ambiguity and high conflict, the coalition strength of the various stakeholders will be the key driver in achieving the outcomes.

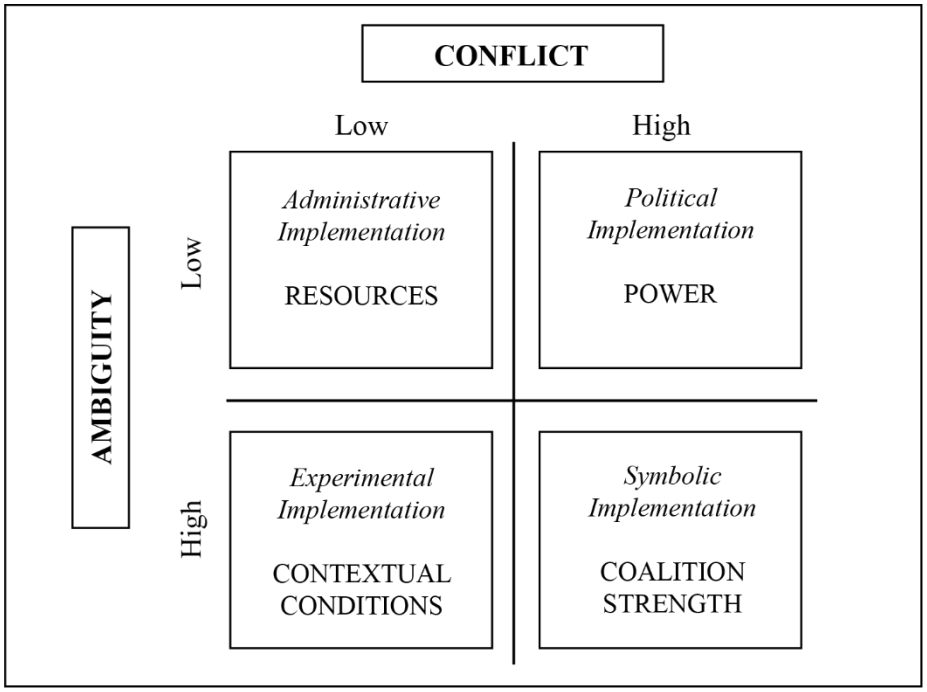


Figure 9: Matland's ambiguity or conflict model

Goggin, Bowman, Lester and O'Toole (1990) developed an integrated communications model, premised on communications theory, and utilised it to explain implementation variations in different social and regulatory policies spanning different levels of government (Winter, 2003). For example, federally prescribed policies will be received very differently from policies where states and local agencies have the freedom to implement their own approaches. Similarly, the interaction between interest groups, state and local level officials and the main agency can either lead to a receptive climate or constrain implementation. State implementation is also affected by the decision of the state to move ahead with implementation, as well as their capacity to implement. The latter does not only refer to the necessary human resources, but also to whether the organisational structure "ease the transmission of information" and whether the financial resources are in place (Goggin et al., 1990, p.38). The final element highlighted in the model is the need for feedback and policy re-design to take place. However, this requires local and state officials to communicate their experiences back to a federal level and secondly for federal officials to distil the feedback from the various state and local agencies and deciding whether to act on it through policy redesign.

Saetren (2014), in assessing the scholarly contributions of the third generation paradigm, reports an increase in quantitative studies since the 1990s, but a less than satisfactory number of studies undertaking cross-national comparisons and longitudinal designs. He concludes that the third generation paradigm has made many methodological contributions as opposed to theoretical contributions and ascribes this to the "demanding nature of the third generation research paradigm" (Saetren, 2014, p. 84).

Interest in the field of implementation research started to decline in the 1990s as evidenced through the reduced research activity in core journals. Towards the late 1980s, both in the UK and USA, right wing politicians came into office and turned to the private sector for solutions in addressing the deepening fiscal crises and rising dissatisfaction of citizens. In the UK, reforms associated with policy analysis took a back seat as all efforts were geared towards regaining control of expenditure (Hill & Hupe, 2014). In the USA, a similar trajectory followed as President Reagan set out to reinvent government by drastically reducing the size of government. This marked a new era in government, referred to as the NPM.

3.3 Implementation under the NPM and NPG

Since the mid-1970s there has been a gradual shift away from in the involvement of government in policy implementation as NPM was introduced into the public sector. Setting aside the country variances in implementing NPM, its application has resulted in government taking on a "steering rather

than rowing" role (Osborne, 2010). Firstly, government services were decentralised to reduce the size of government. In this way the size of the welfare state could be curbed, and costs reduced. Secondly, marketisation as a guiding principle of NPM encouraged competition amongst providers as it is deemed to bring about greater efficiencies. Thirdly, the contracting of outside agencies to undertake service delivery had two results: government officials now fulfilled a predominantly managerial role, and many more agencies became involved in service delivery (often referred to as 'agencification'). Contracting was utilised extensively to govern principal-agent relationships, providing clear guidelines around the desired levels of performance (called performance contracting). Aside from horizontal contracts with contracted agencies, contracts were also being instituted at a vertical level better to govern the relationship between the political level and administrative level. Overall, the NPM is characterised by outside agencies being predominantly responsible for implementation. (O'Flynn, 2007, Osborne, 2010; Verschuere, 2009).

NPM has been criticised for being ill-fitted to cope with the contemporary complexities facing government. The application of private sector principles and "arms-length organisations" led to policy implementation becoming "organisationally distanced from policymakers" (Osborne, 2010, p. 3). It is not surprising that Barrett and Fudge (1981, cited in Schofield & Sausman, 2004, p. 236) view implementation studies to be "most unfashionable" during the NPM phase.

The complexities of societies however "require the expansion of knowledge and a problem solving capacity that cannot be provided by any single entity operating alone" (Rittel and Webber, 1976; Head, 2008 cited in Koppenjan & Koliba, 2013, p.1).

Peters (2014) concurs with the notion that implementation is never undertaken by a single organisation. He distinguishes three types of implementing organisations: firstly, he distinguishes between implementation that only occurs within the public sector, whether it be vertically and horizontally. Peters (2014) makes the point that even within a single government agency, substantial differences exist between higher and lower echelons of the bureaucracy, as well as different levels of governments (e.g. federal, state and local in the case of USA). Horizontal implementation within the government context tends to be particularly complex, given the coordination required to unite different agencies and units around a common goal (Peters, 2014). The second type of implementation structure concerns joint implementation between the public sector and private sector. This is typical to the NPM whereby relationships between market actors are guided by principal-agent theory, to deliver contracted government services. The challenges experienced with this type of implementation structure are two-

fold: firstly, private sector is driven by profit, which could result in high transaction costs. Because of the arms-length relationship, additional monitoring might be required to ensure implementation happens as planned. The final implementing structure is networks consisting of various actors combining forces in pursuit of a common goal or purpose. This can include an array of organisations such as labour unions, third sector organisations and religious organisations (Peters, 2014). The study of networks is a central concern of the NPG; in fact Osborne (2010) distinguishes network governance as one of five strands of work under the NPG.

The NPG came as a response to the shortcomings of the NPM, underpinned by the belief that "working across organisational, jurisdictional and political/administrative boundaries will enable more efficient and effective policy development and implementation" (Christensen, 2012, p. 2). The NPG, with its roots in network and institutional theory recognises the plurality of the public service system. Not only do multiple actors – inside and outside government – contribute to service delivery, but various governance processes need to be taken into consideration (Osborne, 2010).

Some of the key doctrines associated with the NPG that affect implementation are the strong focus on the horizontal dimension of implementation given that wicked problems are typically cross-cutting, extending organisational demarcated boundaries. The NPG is also characterised by a shift from markets to networks whereby a mixture of actors – each contributing skills and resources - assist in the delivery of government services. The NPG re-introduces values such as teambuilding, trust and integration in order to unify government as well as mark a shift away from the command and control approach to policy formulation and implementation, to a greater focus on the procedural elements of policymaking and implementation. Policy formulation and implementation are characterised by cooperation, collaboration, negotiation and active participation as the notion of partnerships move to the fore. This requires public servants to take on the new role as network managers and coordinators of different delivery networks. Under the NPG, attempts are made to strengthen the centre of government thereby returning power to both the administration and political leadership. (Christensen, 2012; Torfing & Triantafillou, 2013).

Much of the NPG approach to implementation is not new. The notion of a stronger centre of government aligns to the top-down theorist way of thinking. Additionally, both first generation theorists, specifically Pressman and Wildavsky and bottom-up theorists (such as Hjern) acknowledge the role of multiple actors and multiple organisations in implementation (Peters, 2014).

The distinguishing factor between the NPG and PA is that whereas policy implementation started out as a sub-discipline of public policy and public administration under the PA it is now "found at the intersection of public administration, organisational theory, public management research and political science studies" (Nilsen et al., 2013, p. 3; Saetren, 2005). The result of the governance and NPM public sector regimes on implementation research are two-fold. Firstly, under these regimes the definition of implementation was significantly broadened to become "a multidisciplinary, multilevel and multifocal exercise looking at the multiplicity of actors, loci and layers" (Hill & Hupe, 2014, p. 17). Secondly, and closely related to the first, is that new topics have been added to the study of policy implementation: "an emphasis to address the effects of institutional and inter-organisational relationships with governance and policy networks emerging as important research topics" (Nilsen et al., 2013, p. 3). The implication of this is that the scope of variables influencing implementation has been broadened (Hill & Hupe, 2014). The need for a continued focus on implementation is therefore more important than ever (Hill & Hupe, 2014).

3.4 Summary

Policy implementation research has a relatively short history compared to performance measurement. Whereas performance measurement is concerned with the whole system of government operations (inputs to outputs to outcomes), policy implementation focusses on the processing phase, i.e. how inputs are processed or transformed to produce outputs, and ultimately results.

Four points emerge from our discussion of policy implementation research in this chapter: i) the varying intensity and location of policy implementation across the three public sector regimes, ii) the theoretical contributions, iii) the methodological contributions and iv) conceptual advancements.

Intensity of policy implementation research

Policy implementation research studies started out as a distinct field in the USA in the 1960s, motivated by the need for a better understanding of why policies failed. The dependency between policy implementation and policy outcomes was finally being recognised. Three generations of policy implementation ensued: first generation research between 1960s-late 1970s, second generation research in the early 1980s and third generation research that covered the latter half of 1980. The majority of first generation research took place during the PA regime. The theoretical underpinnings of this regime significantly influenced the early implementation theorists thinking.

Under the NPM very little attention was afforded to policy implementation due to a reduction in the size of government as well as the use of outside agencies to undertake service delivery. With the NPG implementation gained prominence again: the public servant now becomes a network manager, with policy making happening jointly and implementation transcending organisational boundaries. Unlike the PA regime, policy implementation during the NPG was broadened and taken up under the governance and network banner.

Theoretical contributions

Theory development was a major impetus for both the first- and second generation implementation scholars. The aim was to develop a causal theory that would identify the critical success variables in policy implementation. Within the closed system approach to government, only a limited number of variables were considered by the first generation scholars. This changed significantly with the second generation scholars introducing many more variables into the mix.

Three major theoretical contributions emerge when comparing the first generation scholars with the second generation scholars: firstly, there was an increased awareness of the importance of understanding contextual variables during policy implementation. Despite substantive differences in their approach, top-down and bottom-up second generation scholars agreed that three categories of variables influence implementation: i) policy form and content, ii) organisations and the resources available as well as iii) the people responsible for implementation. Secondly, second generation scholars attempted to empirically test which variables impacted positively on implementation – something which was not done by the first generation scholars. And finally, the bottom up theorists are accredited for raising the importance of networks and involvement of other actors in the implementation process.

The third generation scholars not only tried to narrow down the number of variables but also concluded that a single theory of successful implementation was not feasible. Instead, their focus was on empirically testing the variables identified so far, with the view of developing consolidated and comprehensive models for policy implementation. Under the NPM, implementation was mainly undertaken outside of government. With the onset of the NPG, interest in policy implementation – although not a distinct sub discipline of the policy sciences – resumed. From a theoretical perspective the importance of networks and the involvement of multi actors in policy implementation became a key area of interest under the NPG. In addition, the interest shifted from the traditional hierarchical structure of implementation was replaced with horizontal, multi-level modes of governance.

Methodological contributions

First generation scholars' work can be described as mainly qualitative and a-theoretical in nature. For them implementation failure was associated with either faulty legislation or a lack of compliance to prescribed standards. Even though second generation scholars attempted to empirically test which variables impacted positively on implementation this unfortunately did not result in reduced list of success variables – instead, the number of variables that potentially impacted on implementation increased. Third generation scholars endeavoured to follow a more scientific approach to policy implementation which included experiments and quasi-experiments, longitudinal research design and comparative studies. This unfortunately did not fully materialise, but some useful frameworks were developed that combined the work of the second generation bottom up and top town scholars. Under the NPG regime, the inter-organisational nature of implementation caused for a whole new set of variables around network formation, network interdependencies and network performance to be added into the mix.

Conceptual advancements

The various debates between the top down and bottom up scholars ensured continued advancement in the way policy implementation was undertaken. For example, the top down scholars saw the goal statement as the ultimate evaluative standard as it encapsulated the policy objectives. For the bottom up theorists, the evaluative standard was not the goal statement but rather a local problem statement that must be identified in consultation with the street level bureaucrats who are able to contribute significant "know how". Additionally, the top down scholars' clear distinction between policy formation and implementation was also challenged by the bottom up scholars' more fluid understanding of policy formation compared to policy implementation. This thinking advanced the way in which implementation was approached under the NPG. Under the NPG implementation is regarded as an inter-organisational, as opposed to intra-organisational endeavour. In addition, the complexity of policy problems requires an expansion of participating actors and role players. This places government in the unchartered terrain of enabler instead of controller,

This summary table provides a summary overview of the main contributions from the three generations of scholarly implementation research, and its alignment with the three public sector regimes. Examples of key scholars under every generation are also included

Table 4: Summary of table of scholarly contributions and key underpinnings of three public sector regimes

<table>
<tr><th>Time period</th><th>Public sector regimes</th><th>Policy Implementation research</th></tr>
<tr><td rowspan="4">Late 1890s-1980's</td><td rowspan="3">Classic public administration
- Clear distinction between policy design & implementation
- Decision maker is government
- Government function within a closed, rational system
- Focus on standards as influenced by scientific paradigm</td><td>First generation scholars: 1970s</td></tr>
<tr><td>First generation scholarly contributions
- Attempt to build theory: but focused on policy failure
- Rational, linear approach to theory building: few variables
- Introduced notion of an "implementation deficit" and implementation as "the missing link" in the policy process
Examples of scholars: Pressman & Wildavsky (1973), Eugene Bardach (1977), Erwin Hargrove (1975)</td></tr>
<tr><td>Second generation scholars: late 1970s – late 1980s</td></tr>
<tr><td>Top down scholarly contributions:
- Clear divide between policy design & implementation
- Policy goal is the authoritative decision
- Significant expansion of implementation success variables: ultimately categorised in terms of i) policy content, ii) organisational, iii) people implementing the policy
Examples of scholars: Sabatier (1986), Mazmanian & Sabatier (1989), van Meter and Carl van Horn (1975), Hogwood and Gunn (1984)</td></tr>
<tr><td>Transition period between PA and NPM – not as clear alignment with PA principles as second generation top down scholars</td><td>Bottom up scholarly contributions:
- Shifts attention away from centre of government variables to the contextual variables
- Evaluative standard is not policy goal but a problem statement
- Contribution of actors outside of government recognised
- Blurred lines between policy formation and implementation: policy is also made during implementation
Examples of scholars: Berman (1978), Michael Lipsky (1980), Benny Hjern (1982), Hjern and Hull (1982) Hjern and Porter (1981), Barrett and Fudge (1981)</td></tr>
<tr><td rowspan="2">Late 1980's – 1990s</td><td rowspan="2">NPM:
- Policy implementation largely taking place outside of government</td><td>Third generation scholars: late 1980's-1990s</td></tr>
<tr><td>Third generation scholarly contributions:
- Abandoned idea of a single theory for successful implementation
- More scientific base (quantitative studies) to identify key variables. This includes comparative studies, longitudinal design, statistical analysis
- Development of combined frameworks
- Slowing down of policy implementation research efforts
Examples of scholars: Goggin, Bowman, Lester, O'Toole (1990), Matland (1995)</td></tr>
<tr><td>2000's</td><td>NPG:
- Emphasis on horizontal dimension of implementation (due to problems being cross-cutting)</td><td>Policy implementation not a dedicated scholarly topic anymore – subsumed under banner of public administration, organisational theory, public management research and political science studies</td></tr>
</table>

Time period	Public sector regimes	Policy Implementation research
	- Shift from markets to networks (with network actors partaking in policy formation and implementation) - Return power to center of government as eroded by NPM due to decentralisation and fragmentation of service delivery	

In the next chapter on Deliverology I will delve into what Deliverology is, the Deliverology framework as well as the overlaps with the performance measurement tradition and the policy implementation tradition.

Chapter 4: Deliverology - an approach to accelerate delivery and optimise results

4.1 What is Deliverology?

Deliverology is a reform initiative that originated in the UK during Tony Blair's second term of office. It was developed by Sir Michael Barber, who headed the first delivery unit (called the PMDU). The impetus for a different approach to delivery came from the disappointing results during the UK Labour Government's first term of office. After four years in office, Blair realised that policy implementation was far more difficult than policy development.

> After five years in government, I know only too well that passing legislation or making a speech will not solve vandalism in estates, raise standards in secondary schools, look after the elderly at risk. Indeed the state can sometimes be part of the problem (Blair quoted in Liaison Committee, 2000102, paragraph 37, cited in Smith, Richards, Geddes & Mathers, 2011, p. 980)

Governments are under increasing pressure to deliver. However, a variety of factors impede on government's ability to execute policy: "[P]olitical pressure can cause priorities and resources to shift, success can be difficult to measure, consequences for failed delivery are less obvious than in the private sector, and stakeholder motivation are not always transparent" (Barber, Kihn, & Moffit, 2011, p.32). This leads to "government by spasm", which is characterised by vague aspirations, a focus on the present, an "everything matters" approach and crisis management. The solution is to institute routines that enables government to set clear priorities, identify specific measures of success, be future-focused, have constant dialogues about delivery and base decisions on real-time data (Barber, 2015).

In order to transform government performance, more emphasis needs to be placed on delivery. Even though the concept of delivery is not defined in the Deliverology literature it is often used interchangeably with implementation. Yet, a clear link is made between delivery (implementation) and results with (poor) implementation viewed as the culprit if results are not being achieved. The two definitions of Deliverology below demonstrate the focus on implementation and putting systems in place to address this implementation gap. The Schacter (2016) definition also captures the dependency between implementation and outcomes, while the Delivery Associates definition focuses more on the "how part" of delivery.

> [Deliverology is]an approach to managing and monitoring the implementation of activities that have a significant impact on outcomes (Schacter, 2016, p. 2).

> Deliverology is our system for helping governments deliver meaningful results that will last. It is both a science and an art. The science is the routine of setting a target and then using data, technology, planning, monitoring and problem solving to achieve it. The art is the way you do it and how you behave – it needs focus, urgency, ambition, honesty and humility (Delivery Associates, n.d. cited in Birch & Jacob, 2019).

From a more practical perspective Deliverology is often explained through the following four questions (Barber, Rodriguez, & Artis., 2016, p. 3):

- What are you trying to do?
- How are you planning to do it?
- At any given moment, how will you know whether you are on track to succeed?
- If you are not on track, what are you going to do about it?

These questions are commonly used by Deliverology practitioners to re-direct the thinking around government delivery.

4.2 The evolving Deliverology framework

The UK PMDU, by virtue of being the first delivery unit, shaped the Deliverology approach. Scism (2015) describe the PMDU as utilising "target-based performance management" with the purpose of transforming service delivery by considering both implementation and results. The 11 priorities of the PMDU spanned health, education, the home office and transport – with Richards & Chegus arguing that the success of the PMDU can be attributed to the fact that the objectives "were narrow, departmental-specific, and relatively easy to measure" (2018, p.7).

The main responsibilities of the UK PMDU included: i) assessing and reporting performance of the selected priorities, ii) providing analytical support and recommendations to accelerate delivery capacity and sustain continued improvement, iii) assisting with problem identification and removing obstacles to delivery, iv) supporting the development of policy and the performance management framework for Public Sector Agreements (PSAs) (Scism, 2015). As part of this work, delivery plans were developed

which required departments to specify their strategies for achieving the required PSA objectives. The PMDU developed many tools and delivery routines which included stocktake meetings, priority reviews, delivery reports, monthly notes, delivery trajectories, delivery chains and league tables (Richard & Chegus, 2018). It was already evident from the first delivery unit how important the participation of the head of government is in Deliverology; with a secondary success factor being that the PMDU was located in the centre of government (Scism, 2015). Sir Michael Barber headed the unit for 4 years, after which he left to pursue various career opportunities in the private sector- one being the establishment of an advisory consultancy called Delivery Associates which advises governments on the establishment of delivery units and instituting the Deliverology approach.

A 2011 joint publication with colleagues from McKinsey and Company marks a refinement of the PMDU model into six steps as identified in Figure 10:

- set clear goals,
- establish metrics and lines of accountability,
- set targets, and develop plans,
- track performance effectively,
- hold performance dialogues and
- ensure action is taken.

These denote the typical performance management steps, but with Deliverology these steps are approached differently. The objective setting stage in Deliverology (step 1) requires for prioritisation to be done – this is not typical within government where there is a tendency to tackle everything often without a clear evidentiary base. Step 2 of Deliverology entails the selection of a few key performance indicators with clear lines of responsibility attached to the achievement of these indicators. Given the proliferation of performance indicators within the public sector, this can be challenging to achieve, and require careful consideration of the best indicators by which to track progress. As far as Step 3 is concerned, there is a tendency to err on the side of caution when setting targets given the strong compliance culture in government. Deliverology on the other hand calls for ambitious targets to be developed. Similarly, with step 4 - although a recognised step within the performance management cycle, governments do not always have access to up to date information. The dysfunctional behaviour attached to performance measurement discussed in Chapter 2 also influences the accuracy of data. Deliverology pays significant attention to the availability of data, ensuring access to as real-time data as possible. Robust performance dialogues are not typical to government (Step 5). With Deliverology, effort is made to establish a suitable platform where both the administration and political leaders can

engage around key priorities, on a continuous basis. Step 6, the feedback loop, is a critical step in the performance management cycle. In its typical form, it entails addressing a set of recommendations and implementing the lessons learned. In reality, the persistent follow through, and continuous tracking of the effects of these changes are not always done, resulting in substandard performance. The routines established in Deliverology ensure action is taken, and that consequences exist for good and bad performance.

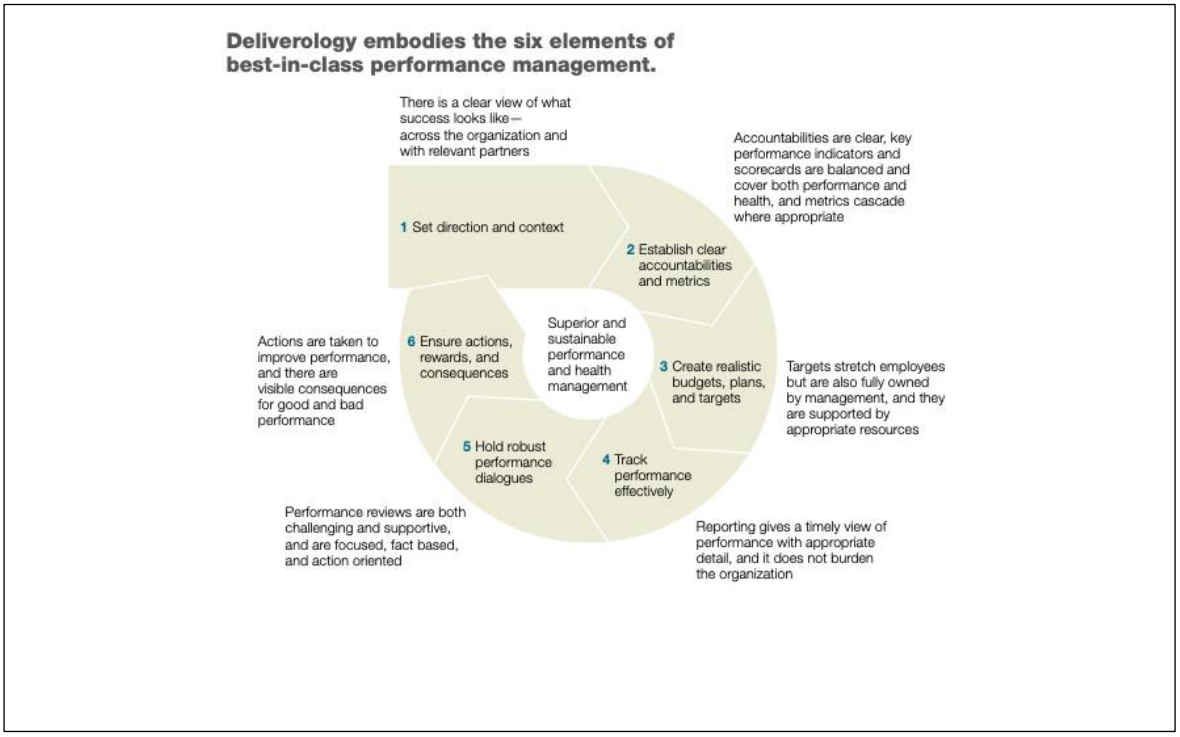


Figure 10: Deliverology steps

(Barber et al., 2011, p. 33)

The steps provided in Figure 10 encompasses the common functions performed by a delivery unit. Table 5 lists the common functions performed by delivery units mapped against the steps shown in Figure 10 as well as aligning it back to the original PMDU model (column 3) It shows that despite some minor differences the functions of performance tracking, target setting and problem solving remained. In terms of problem solving, this is addressed by Step 6 as it requires for action to be taken to improve performance. Step 6 also addresses the issue of accountability, but in practice accountability formed part of step 5 as the routines (such as stocktakes) are utilised to hold officials accountable for specific tasks.

Table 5: Alignment of Deliverology steps to the functions performed by delivery units

Deliverology steps as per Figure 10	Common functions performed by delivery units	Alignment to PMDU model responsibilities
Step 1: Set direction and context	Assist with articulation of clear priorities	Implicit in step 4
Step 2: Establish clear accountabilities and metrics	Develop key metrics and set targets	Implicit in step 4. Also, overall purpose of PMDU seen as being "target based performance management"
Step 3: Create realistic budget, plans and targets	Develop delivery plans	Develop delivery plans (for PSAs)
Step 4: Track performance effectively	Undertake data analysis and simplistic visualisation of data to convey status of performance	Assess and report on performance of the selected priorities Provide analytical support and recommendations to accelerate delivery capacity and sustain continued improvement
Step 5: Hold robust performance dialogues	Establish routines that encourage accountability and facilitate discussions around delivery	The PMDU developed many tools and delivery routines which included stocktake meetings, priority reviews, delivery reports, monthly notes, delivery trajectories, delivery chains and league tables
Step 6: Ensure actions, rewards, and consequences	Identify delivery blockages and challenges, and ensure problem solving	Assist with problem identification and removing obstacles to delivery

(Barber et al., 2011; Gold, 2017; Scism, 2015)

Even within the common functions performed by delivery units, some variations are present when comparing different delivery units. The reason for this is that countries customise the approach to fit their context, leading to modifications and additions to the Deliverology framework. Using the original PMDU as a model, Gold (2017) pinpoints four variations in the way in which some of the common functions of delivery units are performed. The alignment to the steps provided in Table 5 are indicated in brackets):

- the types of priorities being tracked (Step 1),
- how priorities are selected (Step 1),
- the routines being utilised to report on delivery (Step 5), and
- approaches to problem solving (Step 6)

In terms of the types of priorities being tracked, four variations are found: the PMDU dealt mainly with sector specific targets that were the responsibility of one department. As shown in table 6, the variations found in other delivery units are that the priorities tend to be cross-cutting across departments, while some delivery units also track the completion of infrastructure projects, national development plan priorities or only departmental priorities. In identifying the priorities various means aside from the political leader identifying the priority (as was the case with the PMDU) have been undertaken. This includes obtaining input from international donors, concluding collective agreements between countries which contains the agreed priorities or utilising design labs to get different stakeholders together.

Different routines for reporting have also been established by delivery units over time. This includes the development of data dashboards, utilising cabinet meetings to update the political leadership and convening stocktakes (fully-fledged or mini stocktakes). New approaches to problem solving have been developed since the use of "deep dives" by the PMDU. This includes pre-implementation solutions, such as assessing the feasibility of policy proposals to pre-empt delivery challenges as well as utilising existing community structures and design labs to undertake collective problem solving.

Table 6: Key variations in applying Deliverology

Type of priority being tracked	**How priorities are selected**
PMDU Model: key service delivery targets in specific departments	PMDU Model: Prime Minister selected the priorities
• **Cross-cutting outcome targets**: delivery units such as in Canada and the UK tracks outcomes that span across departments • **Completion of major priority projects**: for example infrastructure projects (e.g. in Kenya, Paraguay, and Serbia) • **National development plan initiatives**: examples are Vision 2025 in Jordan, Vision 2030 in Saudi Arabia, and Wawasan Brunei 2035 in Brunei) • **Departmental priorities** as selected by either the department or the head of government (as in Canada and the UK)	• **Input from international donors** (World Bank and EU in Romania) • **Collective agreement** between heads of government and ministers (e.g. Malaysia and Serbia) • **Use of design** labs to obtain input or refine priorities selected by political leadership and heads of government
Routines for reporting	**Approach to solving problems**
PMDU Model: Delivery notes, delivery reports, stocktake meetings and site visits	PMDU Model: deep dives to investigate problems and develop appropriate plans, as well as stocktake meetings to brainstorm solutions
• **Online dashboards** that summarise performance (for e.g. New South Wales in Australia, and Punjab in Pakistan) • **Mini stocktake meetings:** cover lower-order priorities (e.g. New South Wales)	• **Review departmental policy proposals** to assess feasibility of delivery (e.g. Canada and New South Wales) • **Using design labs** as a problem solving space involving delivery experts (e.g. Malaysia)

• **Combination of mini stocktakes and normal stocktakes** (e.g. Canada) • **Cabinet meetings to update political leadership** (e.g. Guatemala and Indonesia)	• **Using cross-government communities of practice** to identify solutions to problems (n Canada there is a monthly meeting of all chief results and delivery officers) • **"Feet on the ground"** to understand problems better and identify solutions

(Shortened from Gold, 2017, pp. 13–14, additions based on Western Cape DSU trip abroad in 2018)

The Deliverology literature show that aside from variations to the approach, additional functions are also being performed by delivery units. The last two bullets of the list below (i.e. data collection and running communication campaigns) constitute the additional functions that the Western Cape DSU performed.

The list of additional functions includes (Gold, 2017; Scism, 2015):

- Providing inputs on policy proposals (both short term advice and long term policy development) for example in Sierra Leone, Rwanda, Malaysia, Australia
- Provide capacity building on all aspects of Deliverology
- Co-design implementation plans
- Explore innovative ways to deliver, as well as solving delivery challenges
- Gather inputs from various entities and communicate results for example in Rwanda, Maryland, Indonesia
- Develop, implement and manage data collection with regard to selected priorities
- Develop and manage communication campaigns in support of the selected priorities

In 2018, the Western Cape DSU undertook a study visit to five delivery units across the UK, North America and Canada with the view of learning about the different approaches and functions performed by other delivery units. The focus during the visit was on Step 1, 4 and 5 of Deliverology, i.e. the type of priority being tracked, the data analysis and data collection as well as routines to track performance. All the delivery units focused on a clear set of priorities; most often the cross-cutting priorities that extend departmental boundaries and which are therefore most challenging to implement. All delivery units visited have routines in place to report on progress – with variance evident in the frequency of stocktake meetings convened and the intensity of coverage. On the data front, most delivery units performed the data analysis and visualisation functions. The data collection function however was undertaken by the relevant government departments.

It was noteworthy that all delivery units performed the additional function of providing formal and informal capacity building, covering the Deliverology approach, the measurement of outcomes and optimising the use of data. This is not surprising given the challenges surrounding outcomes measurement as raised in Chapter 2.

It is clear from this discussion that many countries who adopted the Deliverology approach also made some modification to the framework to align with their own situation: in some instances new or additional functions are taken on, but in other instances the changes only marks a slight variations or modification in the way a step is undertaken.

In part 2 I provide a detailed overview of how the Deliverology framework was modified and expanded for the Western Cape context. In presenting this, the latest version of the Deliverology framework will be utilised to map out the changes and modifications.

The most recent Deliverology framework is found in *Deliverology in Practice* (2016). The framework consists of five steps (Ref. Figure 2) but with a detailed description accompanying each step. It must also be noted that these five steps also contain sub steps which are not shown here (Barber, et al., 2016, p. 5):

1. Develop a foundation for delivery: this includes the development of clear goals and priorities, establishing and capacitating the delivery unit directly under the guardianship of the current political head and build the networks that will assist in driving the transformation.
2. Understand the delivery challenge: analyse existing data and any other information to ensure there is a solid understanding of the current situation and associated problems blocking the achievement of goals.
3. Plan for Delivery: develop a detailed plan, i.e. be specific around what must be implemented, the scale thereof and how it links to what you want to achieve.
4. Drive delivery: Continuously monitor progress against the plan and take corrective action based on what the data and evidence show. Also ensure delivery happens at pace and keeps momentum.
5. Create an irreversible delivery culture: recognise the need for continuous change management to ensure the approach becomes entrenched as part of government culture

4.3 A dedicated delivery unit

To give effect to this much needed emphasis on delivery, a dedicated unit, not too large, with highly capable people is a strong feature of the Deliverology approach. Barber (2015) argues that without this explicit focus on delivery, other functions of government might take preference, notably the policy strategy or policy development functions. The head of the delivery unit reports directly to the political lead and must be a respected, highly capable individual that is preferably appointed on a full-time basis. Delivery unit teams are typically not large and should ideally consist of a mix of private- and public-sector staff. Typical skill sets of delivery unit staff include problem solving, data analysis, relationship management, feedback and coaching, and a "can do" attitude (Barber et al., 2011). The delivery unit should be located outside the hierarchical structures of the bureaucracy to maintain an objective view over the system or organisation it is trying to influence (Barber et al., 2011).

The advantages of having dedicated staff and a delivery function in place are four fold: firstly, it ensures dedicated time and attention is spent on the identified priorities; secondly, without this focus on implementation, policy and strategy tend to take preference; thirdly, delivery units are well placed to support delivery of cross-cutting priorities as they transcend organisational boundaries, and finally, the delivery unit can act as a centre of expertise on implementation and problem-solving techniques, allowing for knowledge transfer and capacity building across the institution (Barber, 2015).

Delivery units are established at all levels of government, but most often at the centre of government. The reason for this is that the centre of government acts as a "support system to the Head of Government as a vehicle for managing the machinery of government" (OECD, 2015, p. 2). The role of the centre of government against the changing landscape of the public sector (titled "emerging issues"), as summarised by Alessandro, Lafuente, and Santiso. (2014, p. 3), is shown in Table 7. Alessandro et al. (2014) show the role of the centre of government in addressing some of the emerging issues facing government. This includes coalescing support around a cross-cutting problem (to break down silos), identify and relentlessly track priority projects (to meet citizen demands for results), navigating the complex, fragmented terrain of unchartered policy terrains and communicating and ensuring a continued focus on key priorities (to mitigate a short term focus).

Table 7: Emerging issues and the role of the centre of government

Emerging issues	Role of the Centre of Government
Most priority issues are multi-dimensional and cross-cutting; they cannot be properly addressed by vertical ministerial silos	Coordinating and brokering solutions, bringing together the relevant ministries and agencies to make decisions, and design and implement policy

Emerging issues	Role of the Centre of Government
Citizens increasingly demand better public services and results from government	Establishing and communicating priority goals (being selective), ensuring budgetary alignment, continuously monitoring progress, unblocking obstacles that affect performance
Government activity has expanded into new policy areas	Supporting and advising the chief executive in managing a complex government structure, especially if policymaking is fragmented or decentralised
A 24/7 cycle, which can deflect the attention of government from priorities	Keeping a systematic focus on strategic priorities, aligning the message from government

Alessandro et al. (2014, p. 3)

As at 2017, several delivery units in the centre of governments had been established across the globe. The figure below reflects the date of establishment as well as the different naming conventions – many delivery units opt not to reflect the wording of "delivery unit" in their title. For example, the Pakistan unit is called the Strategic Support Unit, whereas the Rwandan unit is called the Government Action Coordination Unit.

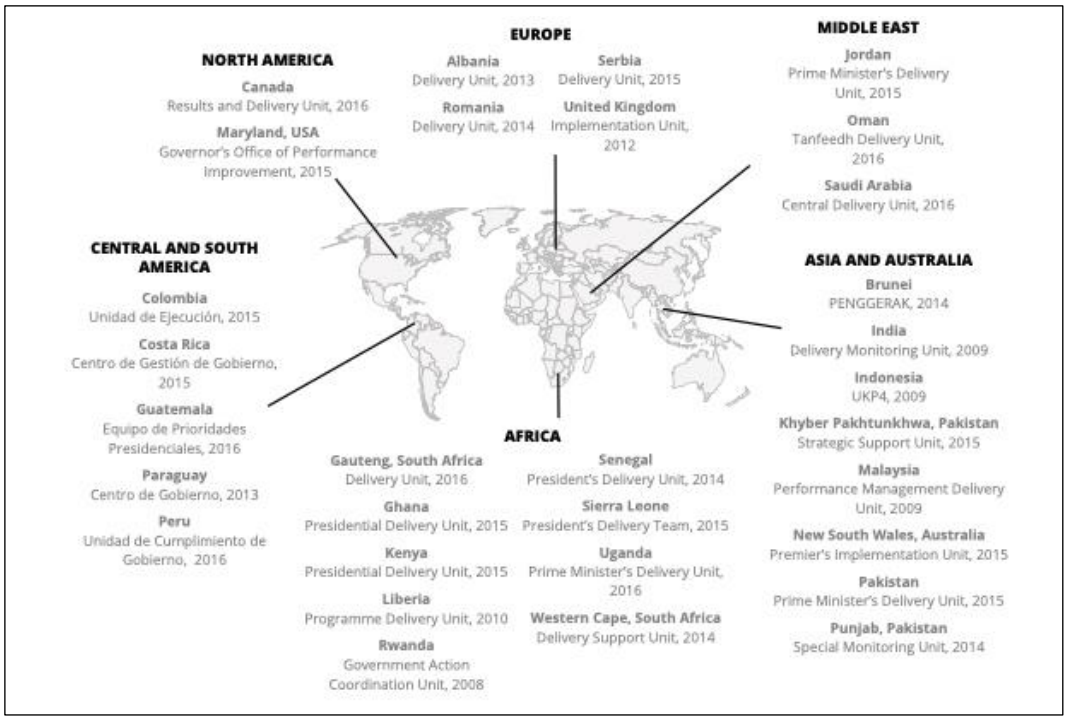


Figure 11: Centre of government Delivery Units (national and regional, dated 2017)

(Gold, 2017, p. 9)

Delivery units are not only found in the centre of government. Other locations include: a) within line departments and agencies (e.g. the Ministry of Public Works in Liberia), as well as b) within local

government as introduced by mayors (e.g. Buenos Aires) and local chief executives (London Borough of Haringey) (Gold, 2017).

Despite variations found in the form, function and locations of delivery units, best practice from across the globe suggests some critical success factors to ensure delivery units do not become DINOS (delivery units in name only) (Barber, 2018). These factors include having strong, highly visible political backing for the delivery unit as well as ensuring the delivery unit is in close proximity to the political head. It is important that a delivery unit only commit to a few strategic priorities, and that these priorities reflect the things that matter to the citizens. As far as possible, cross-government ownership and commitment should be obtained for the selected priorities. Another key success factor of a delivery unit is the institution of effective performance management routines. Data is needed to not only drive delivery, but to ensure the desired results are being achieved. It is also imperative that the delivery unit understand the delivery system in order to provide credible and relevant support. This links to the need to involve stakeholders in all aspects of delivery, including when problem solving is being done. (Gold, 2017; Todd et al., 2014; Scism, 2015).

In the next section I discuss the commonalities between Deliverology, the performance measurement tradition (Chapter 2) and policy implementation (Chapter 3).

4.4 Performance measurement and policy implementation: roots and links with Deliverology

With Deliverology commencing during the NPG, it is not surprising that the approach exhibits key characteristics associated with this regime. Deliverology also has strong roots in the NPM, as well as other performance related reforms in the UK and the USA. NPM's significant influence on Delivery is evident through its managerial approach, its strong emphasis on performance measurement and the need to achieve results. The quotes from Richard and Chegus as well as Prime Minister Blair demonstrate the NPM's theoretical underpinnings:

> Deliverology employs goal setting, performance measurement, and the use of tight feedback loops to ensure that departments deliver on their campaign promises. In this respect, it is consistent with other results management approaches noted in public management, and in fact was a component of broader management reforms in the UK

> that include the use of performance agreements and business plans. (Richard & Chegus, 2018, p. 1).

> [I]n domestic policy, changing public service systems inevitably meant getting into the details of delivery and performance management in a radically more granular way. Increasingly, prime ministers are like CEOs or chairmen of major companies. They have to set a policy direction; they have to see it followed; they have to get data on whether it is; they have to measure outcomes (Blair, cited in Barber, 2015, p. 12).

The NPG's influence on Deliverology is also apparent: reforms such as "Joined up Government and "Centre of Government" initiatives were undertaken after the NPM to address the diminishing lack of control at central government level, as well as addressing the "structural inefficiencies" of decentralisation by encouraging greater collaboration and integration of entities (Scism, 2015). Christensen (2012, p.3) specifically cite the introduction of "inter-agency, collaborative units" to better coordinate cross sector and whole of government activities. Delivery units are one example of this.

Another influencing factor was the emergence of evidence-based policymaking (EBPM) as part of the UK public sector Modernisation Programme. The "Modernising Government" White Paper (March 1999) marked the starting point for evidence-based policymaking in the UK and was followed by the establishment of a number of new units to advance EBPM within the UK government (Parsons, 2002). Also included in the Modernisation Programme, were the development of PSAs and an accompanying performance measurement framework. The PSAs required that targets be developed for specific areas of government (Richard & Chegus, 2018). The PMDU was tasked to assist with the development of the PSA policy and performance measurement framework (Scism, 2015).

And finally, the expansion of data driven governance in the USA through the "stat model" as introduced by the New York Police Department also inspired the way data was utilised to problem solve under the Deliverology approach. Martin O'Malley, first as Mayor of Baltimore and later as governor of Maryland, utilised this data driven approach before Deliverology was conceptualised (called CitiStat and StateStat respectively) (Scism, 2015).

Manning and Watkins (2013) agree that Deliverology with its strong emphasis on results reside within the results based management sphere of approaches. However, the focus of Deliverology on problem identification and resolution, leads to a sub-categorisation of what Manning and Watkins coin the

TRDM approaches: “targeting results, diagnosing the means”. Manning and Watkins (2013) argue that the TRDM approaches mark a shift from previous technical reforms that focused on “what the public sector looks like” to understanding what is not being achieved and how this can be remedied.

In terms of policy implementation, there are several linkages between the first and second generation policy implementation scholars and Deliverology. The first generation policy implementation scholars were responsible for identifying the “implementation” gap and the fact that without focused attention on implementation the results will not follow. The entire Deliverology approach rests on this principle: without a focus on implementation the results will not happen. In terms of the second generation top down scholars, three links can be made: firstly, the second generation top down scholars made a clear distinction between the policy formation and policy implementation phase. This aligns with Deliverology whereby it is premised that the imbalance between policy formation and policy implementation needs to be addressed. This implies that Deliverology, like the second generation top down scholars view policy formation and policy implementation as two separate steps. Another overlap between Deliverology and the second generation top down scholars, are the fact that the policy decision constitutes the evaluative standard against which implementation success or failure are assessed. Decision making at the centre of government is therefore a characteristic of both Deliverology and the second generation top down scholars. The other contribution from the policy implementation literature is the emphasis on outside actors and their role in policy implementation. Both the bottom up second generation scholars as well as implementation under the NPG recognises the importance of involving outside stakeholders. With Deliverology, external stakeholders are involved in the planning and implementation phases and the development of the delivery chain tool is utilised to map the various stakeholders’ contribution.

First and second generation scholars set out to develop a theory of successful policy, in an attempt to identify the critical success variables. This did not materialise and the third generation scholars focused their efforts on synthesising the work of the first two generations of policy implementation scholars. What transpired from the theory development efforts are the categorisation of success variables along three lines: the policy form and content, the people involved in implementation and the organisational context and resources available for implementation. This is in alignment with the success factors identified by Deliverology: the political backing and strong delivery teams are the backbone of the approach. In addition, the complexity of the policy and content will determine the delivery efforts and resources required to execute the few selected priorities. And finally, the organisational setting must be conducive to delivery: routines, based on real time data must receive priority.

4.5 Does the Deliverology approach produce results?

There is a growing body of knowledge on Deliverology, mainly through the production of case study reports. The available Deliverology literature tends to focus on delivery units with a longer history (i.e. UK PMDU, Malaysian Performance Management And Delivery Unit (PEMANDU) and the Special Monitoring Unit in Punjab), but there is also an expanding number of developing country case studies being produced (for example, for Chile and Latin America as a whole).

On the positive side, Deliverology provides a strong focus on the strategic priorities of government, engages those responsible for delivery to problem-solve and implement solutions, and enhances accountability (Alessandro et al., 2014; Richards & Chegus, 2018). The Chilean government case study says that one of the main contributions of Deliverology is that it turns vague electoral promises into cross-cutting, results-based action plans:

> The delivery management methodology posits the need to orient a government program towards achieving results for the public benefit that can be objectively measured, with time frames and persons responsible, and that are actionable based on a plan that contains a significant element of intersectorality ... This is important, given that actions are normally defined within each sector without necessarily taking into account the impact they might have on other areas of government, or assuming that other sectors will continue to carry out certain tasks, which is not always the case (Alessandro et al. 2014, p. 71).

Deliverology also accelerates delivery due to the high level attention it receives:

> [D]eliverology, given its link to political power, generally brings with it authority, resources, flexibility and a striving for provision of timely advice and quick turnaround (i.e. a sense of urgency that can potentially cut through bureaucratic roadblocks to action) (Birch & Jacob, 2019, p. 310).

A literature review of the PMDU in the UK also shows that officials valued the opportunity to engage directly with the prime minister and found the collaborative problem solving approach, called “deep dives” very useful (Richards & Chegus, 2018).

When considering the achievement of outcomes (and its quantitative targets), I focus on instances where national achievements have been reviewed by other scholars or sectoral experts. As can be seen from

the UK, Malaysian and Punjab examples below, delivery units do produce results, but these results are often questioned on the basis of data validity, the contribution or role of the delivery unit in achieving the targets and target setting practices.

In the UK healthcare service, hospital waiting lists dropped from 40 000 in 2001 to fewer than 10 000 in 2003. This needs to be viewed in the context of a 30% increase in public expenditure on healthcare during Blair's term of office (Hood, 2006; Manning & Watkins, 2013). Some discrepancies were also found between the official data reported at national level, compared to that collected by means of customer surveys (Bevan & Hood, 2006). The heavy emphasis on targets also had two other unintended consequences: the waiting time for patients with less serious injuries increased, and patients tended to be discharged too early, leading to relapses and even death (Reevely, 2016, cited in Richard & Chegus, 2018; Seddon, 2009, cited in Richard & Chegus, 2018).

In the UK education sector, instead of advancing under the Deliverology approach, education took a step backwards: teachers were forced to use rigid lessons plans and "questionable teaching practices" (Richard & Chegus, 2018, p. 12). An evaluation by Coffield (2012, cited in Richard & Chegus, 2018) also shows bigger improvements in test scores prior to the Deliverology approach being implemented.

THE UK street crime initiative (SCI), aimed at reducing robberies and other street crimes, had great success: robberies dropped by 32 % between 2002 and 2005. However, following its selection as a priority, the SCI received vast resources, the maximum allowed within the terms set by the comprehensive spending review. The SCI also received constant attention from the prime minister, begging the question whether it was not the additional attention and budget that made the difference (Smith et al., 2011).

PEMANDU's priorities included: reducing crime and corruption, improving learning outcomes, reducing poverty and improving rural infrastructure and urban transport. Two years into operation, drastic improvements in crime statistics were reported: crime was down by 11%, exceeding the target of 5%. Improvements were also seen in the corruption perception index (Manning & Watkins, 2013).

An international review panel provided a positive assessment of the PEMANDU results, but some concerns were raised by McCourt (2012) about the lack of evidence on rural infrastructure targets, partial reporting on the corruption result area, discrepancies in reporting on the achievements and the

lack of follow through in auditing the reported targets. The vast drops in crime statistics were also questioned as this was not aligned to the experience of citizens on the ground. The problem relates to the manner in which crime statistics are compiled, with Malaysians feeling that "police were not transparent about how crime statistics were compiled, and suspicions that certain incidents were underreported" (World Bank, 2017).

The delivery unit in Punjab geared at educational reform (called the Programme Monitoring and Implementation Unit) reported success on a variety of their educational indicators. These include enrolling an additional 1.5 million children in school, increasing student and teacher attendance, hiring new teachers on merit and ensuring access to new textbooks for every student and easy-to-use lesson plans for every teacher. Two independent studies were undertaken during the implementation of the Punjab Education reform roadmap: the Independent Commission on Aid Impact (ICAI) and the Department for International Development (DfID) initiated evaluation undertaken by Semiotics (Barber, 2013). Both evaluations confirmed the quality of the education roadmap, and the positive effects this approach had in Pakistan. The data collected through the Annual Status of Education Report (ASER), released in January 2013, confirmed most of the data collected by the delivery unit themselves, with some discrepancy in the teacher absenteeism data. The ASER report did however agree with the i) a significant increase in student attendance data from 81% to 86%, and ii) a 17% increase in schools with toilets (up from 70% to 87%) (Barber, 2013). Jishnu Das, a research economist at the World Bank, and education expert, questions these findings, arguing that a positive trend was already evident in some of the data being studied prior to the involvement of the delivery unit. The credibility of some of the data was also questioned, leading to Das concluding "The lack of credible public data and the lack of third party evaluations of the program make it difficult to go to bat for the deliverologists at this time" (Das, cited in Schacter, 2016, p. 8).

Aside from the country specific examples, another performance-related criticism of Deliverology entails the fact that another reporting layer is added, in an already report-weary environment (Watkins, 2013). This criticism must be viewed within the context of the UK at the time Deliverology commenced. The uptake of performance measurement practices in the UK was so extensive that Hood (2006) describes it as "targetworld mid 2000s UK style" and constitute a broader problem of the performance measurement tradition that is not attributed only to Deliverology.

> The UK Labour Government introduced more than 300 headline performance targets applying to all government departments in 1998 … Each of the headline targets negotiated with the Treasury was accompanied by a larger set of performance indicators, and central

> government departments, in turn set more detailed targets for the delivery organisations for which they were responsible (Hood, 2006, p. 515).

In addition, when measures imposed from central government or head office do not align to officials' own performance measures, a compliance mentality kicks in that counteracts the efforts of a result-based approach. (Bevan & Hood, 2006). Another general point of criticism against the Deliverology approach is the fact that it can potentially undermine other government accountability efforts, such as attempts to institutionalise M&E (Birch and Jacob, 2019).

Measuring results, as seen from Chapter 2 has challenges. This is not different for Deliverology. Firstly, establishing the impact of the delivery units in terms of achieving the desired results are complex. On one hand, the results achieved by some of these delivery units are still very recent and the time frames for achieving the results have yet to be met. On the other hand, as stated by Behn (2017), Gold (2017), and Scism (2015), it is difficult to measure the added value of delivery units, because they usually do not have internal mechanisms to capture their impact (such as formal impact evaluations), they tend to be part of a more comprehensive public policy and administrative reform, and they are careful not to take credit for the results obtained by ministries or entities (Lafuente & Gonzalez, 2018).

Secondly, the selection of the right indicators is also highlighted by PEMANDU: "If the indicator does not measure the desired outcome, then the Ministries, Departments and Agencies meet the target but miss the point" (World Bank, 2017, p. 13).

With this in mind, I recall Birch and Jacob's notion that Deliverology and Evaluation constitute a "Tale of Two worlds" given Deliverology's limited grounding in "scientific thinking" and scientific practice (2019, p.318). It is however the premise of the researchers that instead of divergent paths –Programme Evaluation can address the outcome measurement challenges as well as assist in establishing causality through the development of sound causal theories. This speaks directly to Loughland and Thompson's (2016, p 125, cited in Richard & Chegus, 2018, p. 12), concern that with the Australian education system a handful of indicators are inadequate to capture the complexity of some of the challenges facing government:

> [A] form of epistemological reductionism, that cannot account for the multiple purposes of education, the possibility of intelligent problem-solving by professionals in specific

contexts, and the limitations that deciding what works best imposes on identifying problems and their solutions.

The development of a casual theory ensures the development and selection of indicators that align to the achievement of outcomes. This scientific approach to measurement enables for a more holistic picture of performance to emerge, which could benefit the Deliverology approach.

Delivery units have varying levels of success and longevity. The reason for this is that a delivery unit is established with a specific purpose in mind. Gold (2017, p. 10) makes the points that "alongside the proliferation of new units is a parallel trend in older units closing". This is not surprising, given the highly political nature of delivery units. As new political parties come into office, new reforms are often introduced, resulting in delivery units losing the indispensable political leadership backing needed to operate efficiently. However, in some instances the model has persisted and has even been reintroduced following a period of disbandment. The UK is a prime example. Following the closure of the PMDU in 2010, an Implementation Unit was established in 2012 again to oversee the implementation of government priorities. The map below (Figure 12) provides a snapshot of the delivery units that have been abolished (red dots), units abolished but then re-introduced (green dots) and where the unit has been repurposed with the delivery function falling away.

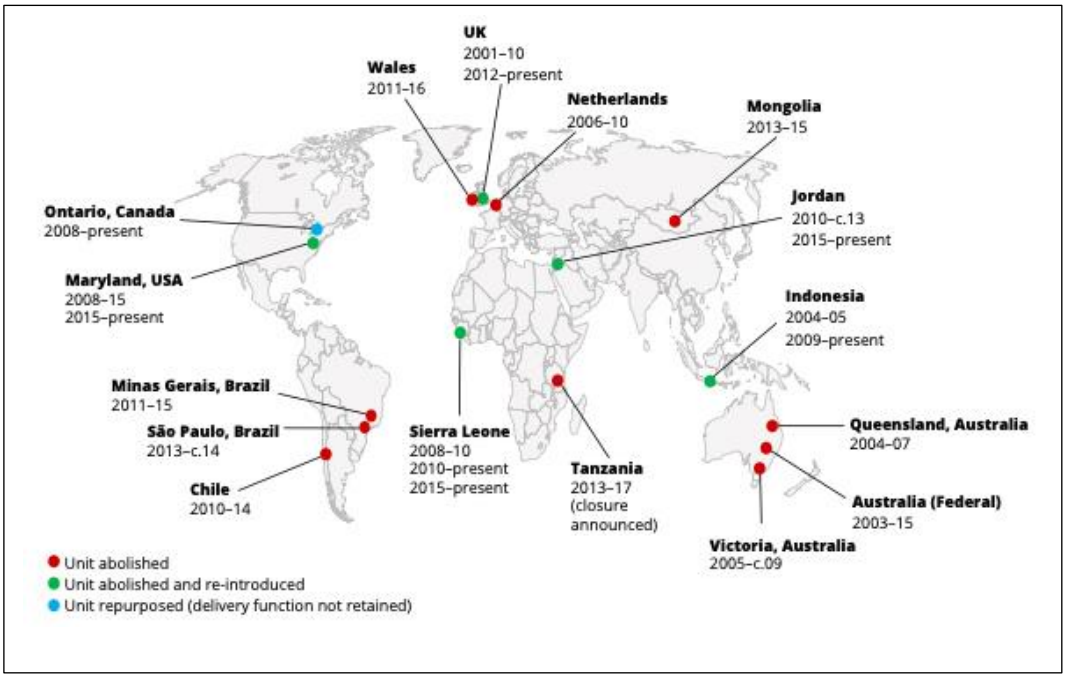


Figure 12: Location of past delivery units

(Gold, 2017, p.11)

4.6 Deliverology as institutionalised in the Western Cape DSU

The first provincial delivery unit in South Africa was established in the Western Cape province in 2015 – called the DSU. The DSU's application of Deliverology found expression by way of the 4 "P" model: the importance of setting **p**riorities, getting the right **p**eople involved, continuously tracking the **p**erformance of the Game Changers and rapid identification and resolution of **p**roblems.

Prioritisation

The starting point was the broad policy priority areas identified by the provincial cabinet. The newly established DSU was tasked to narrow the focus of the Game Changers and to develop detailed delivery plans. At that point only broad statements of intent were formulated by the provincial leadership, i.e. reduce alcohol harms, skills development, energy security etc. Significant work was needed to move these broad policy statements into focused strategies and implementation. The DSU went about this in two ways: firstly, by means of what we referred to as a design lab process, followed by further facilitated sessions to develop detailed outcomes and delivery plans. The design lab process was typically facilitated by external consultants[7], whereas the additional facilitated sessions were run by the performance tracking manager in the DSU. Four of the six priorities followed this combined approach: energy security, alcohol harms reduction, after school and better living model (which was a mixed use, mixed property development).

All the Game Changers were transversal in nature (i.e. multiple departments are responsible for realising the stated goal and accompanying outcomes) with some having the added complexity of requiring the collaboration of external partners and other spheres of government. Three of the Game Changers were geared towards the youth (eLearning, After School and Apprenticeships), with Apprenticeship also supporting the economic growth priority along with the Energy Security and another priority, project Khulisa (which was a departmentally-run programme that prioritised the economy). A further two Game Changers are aimed at redressing some of the societal ills brought about by the apartheid era (Alcohol Harms Reduction and the Better Living Model). The Broadband Game Changer had been running since

[7] The design lab has its origin in similar public sector delivery related work by McKinsey consultancy, which required the involvement of leaders and high-level staff in planning process, sometimes 6 weeks at a time. This approach is intended to ensure the commitment of all stakeholders and is viewed as the critical first step towards successful delivery. The WCG followed a mini design lab process that entailed a two-day lab for each Game Changer, with syndicate groups appointed to do further research, followed by participants reconvening for the final two-day labs.

the previous term of office under the direction of the administrative head of government and hence did not fall under the purview of the DSU. However, its implementation was closely monitored as part of the eLearning Game Changer, given the interdependency of these two programmes. Project khulisa was implemented directly by the Ministry of Economic Opportunities and was never formally adopted as a Game Changer.

Getting the right people involved

Significant effort went into putting the Game Changer teams in place – both in establishing the DSU but also in selecting people for the departmental Game Changer teams. The DSU consisted of two branches of work (see figure below): every Game Changer had a lead and in some instances an analyst to support the work as well as staff that worked across Game Changers, consisting of a communication director and the performance measurement team. This mix of Game Changer specific specialists and cross-cutting focus was not initially planned for in the organisational structure but emerged as the most optimal arrangement as the DSU work got underway. Many of the Game Changer leads got quite close to delivery and having the data function separate within the DSU assisted in the DSU maintaining its objectivity in terms of performing the data analysis and reporting functions. Another benefit of a separate data team was that it could deal holistically with the Game Changer data requirements.

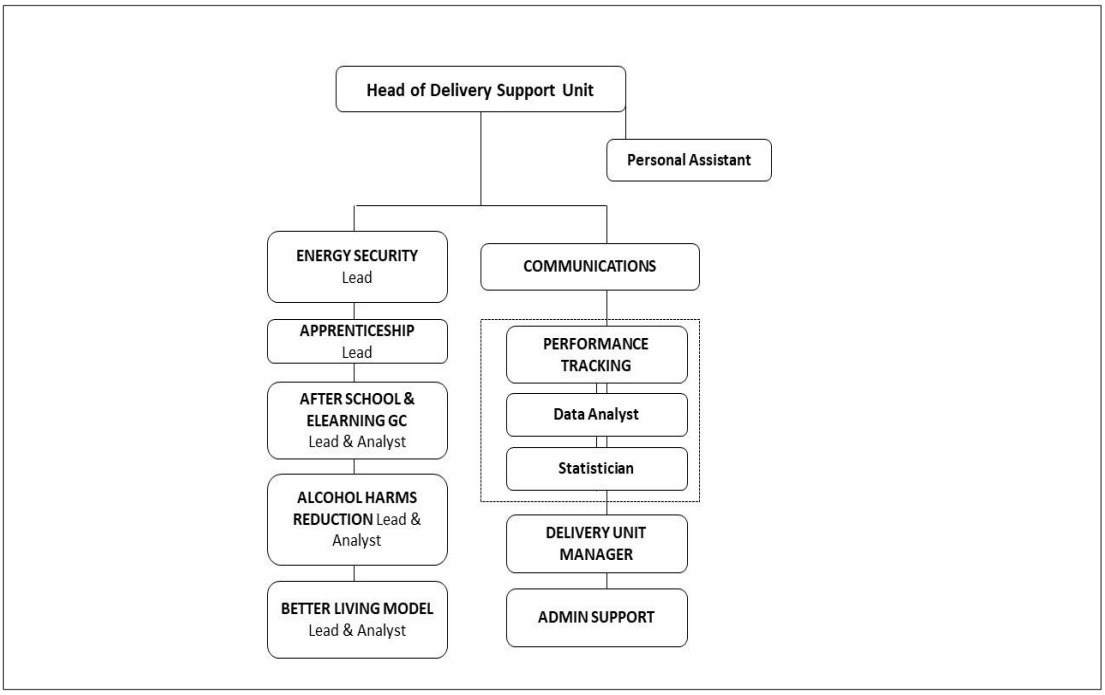


Figure 13: Staffing structure of DSU

The cross-cutting nature of most of the Game Changers meant that multiple departments had to work together in a coordinated manner. Organisationally, a multi-tiered governance structure was put in place to support this transversal way of working. This entailed having a lead department and support department(s) in place for every Game Changer. The lead department took overall responsibility for the implementation of the Game Changer, which meant a) putting in place a dedicated operational team in the lead department and b) establishing a management committee that consisted of the departmental leaders involved in the particular Game Changer. At an operational level, the support departments identified a Game Changer liaison/ lead person that would keep close contact with the Game Changer lead of the operational team. A management committee (MANCO) typically chaired by the lead department's head of department (HoD) provided the first level of executive oversight of the performance of the Game Changer. The MANCO included representation from heads of support departments and senior management officials. This structure ensured clear lines of responsibility between the lead departments and support departments.

The highest level of governance was the two-monthly stocktake meetings, chaired by the Premier and attended by the relevant provincial minister (s) and senior officials (HoDs and departmental staff). The stocktakes provided the focal point for accountability and problem-solving / clearing of blockages, and was the main routine employed by the DSU to track the performance of the Game Changers.

Measuring the performance of the Game Changers

Since I elaborate in the remainder of the thesis on the performance measurement of the Game Changers, only a few points will be raised here. Firstly, detailed delivery plans needed to be developed for all Game Changers against which performance could be tracked. It was during this stage that some limitations to the Deliverology framework were identified and which lead to the incorporation of some principles and concepts from mainstream programme evaluation in the design of the Game Changer delivery plans (called roadmaps). This also entailed the development of indicators, as well as target setting.

Secondly, it became evident early on that the government departments lacked the capacity to collect the required data. The DSU performance tracking team was therefore not only involved in developing the indicators, but also ensuring the execution of most performance tracking activities.

Finally, the institutionalisation of routines facilitated frank and honest conversations around the status of the Game Changer programmes. Stocktakes were scheduled approximately every 2 months for the

duration of the Game Changer programme and were chaired by the Premier. Participants included the relevant political and administrative leadership, Game Changer team members and DSU staff. Significant effort went into preparing the stocktake reports. The stocktake routine was invaluable in ensuring flexibility and responsiveness to the problems at hand. It facilitated transparency and collective thinking around how to unblock the problems which significantly accelerated delivery.

Rapidly identify and solve problems

Given the challenges faced in delivery, problem solving is a key value add of a delivery unit. The dominant compliance culture in the South African public sector restricts an innovative, learning culture that can tackle delivery problems head on. Some of the challenges faced in the public sector include the lack of granular, up-to-date and readily available data to pinpoint the point of breakdown. Decision-making should be evidence based as opposed to anecdotal. Assumptions must be challenged by sound data that is provided timeously to ensure the rapid identification of problems. In addition, officials do not tend to leave the confines of their office to engage with those impacted by their work. Although data is useful it only provides a partial view of reality. Getting out in the field to obtain a fuller understanding of the challenges being faced, enabled more appropriate solutions. And finally, governments' set budget cycles and planning processes do not accommodate the realities of implementation and the need to deviate if necessary. Under the delivery approach, if evidence shows that a project or plan is not working, there is an openness to changing track. This kind of flexibility is not always easy in a bureaucratic environment where budgets are pre-approved and annual performance plans are fixed. (DSU, 2019a)

Another strategy successfully employed by the DSU to problem was getting feet on the ground and visiting the communities and stakeholders being targeted by the various projects and programmes. The Premier herself also supported this approach and started visiting schools on a weekly basis to become more in touch with their realities and accurately identify the persistent problems to delivery.

There were many successes in applying the Deliverology approach in the Western Cape.

A major advantage of Deliverology, is the dedicated focus it provided in terms of delivery. There are many distractions in government, and this approach enabled the teams to focus on the selected priorities, and to delve deep when blockages occur. Once the six Game Changers were approved, the DSU - as recommended by the Deliverology approach - did not deviate from the selected priorities, neither did it take on any additional priorities. This dedicated attention to a select number of priorities is viewed as a "make or break" factor for a delivery unit.

On the people front, the single biggest success factor to the Western Cape DSU was the Premier's unwavering commitment to the Game Changers. Aside from her consistent attendance at stocktakes, she fully engaged with the detail of every Game Changers. However, commitment differed in terms of ministers and top management support for the Game Changers. Where the relevant minister and top management team supported the approach, delivery got underway quickly and momentum was maintained. Where this support was lacking, Game Changers battled to get traction and keep the necessary pace over the three years. As suggested by Deliverology, the DSU was located in close proximity to the Premier (on the same floor). The head of the DSU had direct access to the Premier; a weekly meeting was scheduled between the Premier and head of the DSU. The head of the DSU was also able to call on the Premier at any time should her intervention be required. The DSU staff consisted of a mix of government officials and personnel recruited from outside of government. This staffing structure offered a combination of public sector "know how", proactiveness, pace of work, creative problem solving and sector expertise (for example, the alcohol Game Changer lead who worked for a prominent private sector company in the alcohol industry). This diversity proved to be a winning combination.

Efforts were made to align the performance management systems (i.e. indicators developed as part of Game Changer plans) with departmental performance plans, called annual performance plans, so officials would not feel overburdened, resentful or non-responsive to the Game Changer performance indicators. This was however mostly done for output indicators, not outcome indicators as outcomes are typically cross-cutting, and more challenging to influence.

The independence of the DSU was a major advantage – even though the DSU was perceived to be a politically supported unit, it was still viewed to be part of Government. This arms-length relationship between the DSU and implementing departments meant that the DSU maintained its independence and objectivity, as it was not directly responsible for the delivery of the Game Changers. The stocktake routines significantly improved accountability. Every milestone and/or output was attached to a specific official, not an organisation. This meant direct accountability for achievement/ non-achievement of milestones and outputs.

Another advantage of the stocktakes was the quick identification of problems and arriving at suitable solutions. This improved responsiveness to problems, as some blockages could only be addressed by the Premier and ministers. While not mentioned explicitly as part of the Deliverology approach, an added benefit of the stocktakes was risk management. The fact that the stocktakes facilitated a collectively agreed upon, transparent decision on a difficult matter provided comfort to officials that they would not be left to "carry the can" in the event of things going wrong.

There were also limitations and lessons learned. In selecting the priorities, one must consider the level of control or influence over the results that are being sought. Three of the Game Changers required extensive local government and private sector involvement and support. This made it challenging to drive implementation from the provincial government level. The DSU faced significant resistance at the start. Officials found the approach invasive and did not see the value of another reporting layer. As the stocktake routines settled and data became available consistently, officials started seeing the value of having timeous evidence to base decisions on. Also, having the undivided attention of the Premier and minister accelerated delivery and helped resolve some of the persistent blockages.

We also learnt that a change management programme should accompany the Deliverology approach as significant behavioural change is required to implement Deliverology. This was identified as a critical gap of the "standard" Deliverology approach. Unfortunately, the DSU had not planned for this in its initial staffing, and once underway, it had neither a position nor the budget available. In hindsight more attention should have been paid to the internal communication about the Game Changer programmes to ensure it reached the lower echelons of government where implementation happens.

Significant capacity building had to be undertaken – this ranged from basic project management skills to data analysis and reporting capabilities. Training was also conducted on Deliverology tools and principles in order for departmental teams to become accustomed to the approach.

Despite these limitations, good results were achieved across the Game Changer programmes. The table below excludes the eLearning Game Changer as its performance will be covered in part two of this study. Column 2 provides the target(s) for the Game Changers, while the third column shows the achievement of the targets. Stretch targets were set in most instances, although with some Game Changers it was later established that the baseline was incorrect (e.g. After School Game Changer). Also, with the Apprenticeship Game Changer, the data was not fully up to date by the time the last stocktake happened, hence the slight under achievement. We did not set numeric targets for the Alcohol Game Changer in 2015. This Game Changer was approached as a "test kitchen" approach with different solutions tested in different settings. The results shown in the May 2019 column were just some of the high-level findings from these various initiatives. Despite all the targets not being fully achieved, the good progress that was made can be directly attributed to the leadership commitment, having dedicated Game Changer teams (in the DSU and in departments), continuous performance tracking and the speedy resolution of problems.

Table 8: Goal level achievement of Game Changers: 2015 compared to 2019

Game Changer	Goal statement	Target: 2015	Results: May 2019
After school programme	To ensure regular and sustained participation in after school activities	112 000 learners attending quality after school programmes	• Increased number of learners attending quality after school programmes by 182% (from 29 000 to 81 180)
Apprenticeship	To achieve sufficient, appropriately qualified technical & vocational skilled people to meet the needs of prioritised economic growth areas in the Western Cape by 2019	• 13 221 gain access to apprenticeship programmes • 11 037 apprentices entering labour market	• 11 275 learners entering workplace-based training opportunities • 8 443 qualified apprentices entering labour market
Energy security	To ensure energy security that supports economic growth in the Western Cape, incorporating diverse and low carbon sources of energy by 2020	• 135 MW from small scale solar panels by 2019 • Reduce current Western Cape demand from Eskom by 10% or 260 MW	• 600% increase in uptake of Rooftop Photovoltaic (from 18 MW to 112.2 MW) • Reduced demand from Eskom by 1.66%
Alcohol harms reduction	To reduce alcohol related harms in targeted areas in the Western Cape	• Strong partnerships with communities and law enforcement to reduce alcohol harms • Alternatives both economic and recreational to limit attraction to alcohol	• Sense of safety increased by 21% in Town 2 in Khayelitsha • Increase in compliance of liquor outlets in Game Changer areas, with fines of R828 000 issued between February 2017 and December 2018
Conradie better living model	To break ground on the Conradie site by July 2018	• Break ground on Conradie site by July 2018	Plans, rezoning and appointment of developer delivered in record time: sod turn and handover of site to developer in Jan 2019

(DSU, 2019a)

4.7 Learnings from Part One

Deliverology is a recent approach geared at policy implementation, starting in the UK in the early 2000s. Deliverology has since been applied globally and in 2015, the first South African based delivery unit – the DSU - was established in the Western Cape. This chapter provided a brief overview of Deliverology under five headings: i) the rationale for Deliverology, ii) its theoretical underpinnings and methodological underpinnings iii) the evolving Deliverology framework and the functions of a delivery unit, iv) the effectiveness of delivery units and v) the DSU's application of Deliverology.

Rationale for Deliverology

Deliverology is first and foremost concerned with the delivery (or implementation) shortfalls in government. A disproportionate amount of time is spent developing policy, compared to implementing the policy. This is attributed to many reasons but in the main government with its vast responsibilities have become inept in dealing with delivery. In line with the scholarly literature on policy implementation (in particular the first generation policy implementation scholars), Deliverology recognises that in the absence of implementation, results will not be achieved.

Roots and underpinnings of Deliverology

Deliverology has roots and linkages with both the performance measurement tradition and policy implementation traditions. In terms of global influences, I showed that Deliverology has strong roots in the NPM given its emphasis on performance management principles, managerialism and results-based management. With Deliverology starting during the NPG it also aligns with the "joined up" government and centre of government reforms which sought to rectify the fragmentation and dispersed control caused by the NPM. Deliverology was also influenced by UK-based initiatives such as the Modernisation programme, which not only introduced evidence based policy making into the UK public sector but also the development of Public Sector Agreements that required government agencies to set targets against their objectives.

On the policy implementation front, Deliverology display significant alignment with the second generation scholars. Similar to the top down theorists, Deliverology's starting point is the policy directive which constitute the standard against which performance is evaluated. A second point of alignment between the top down theorists and Deliverology is the clear distinction between policy formation and policy implementation in Deliverology: the premise is that too much time is spent on policy formation as opposed to policy implementation, implying agreement that these constitute distinct activities. The importance of networks and inter-organisational collaboration was raised first by the bottom-up second generation scholars, and later expanded on through the NPG. Deliverology also recognises the critical role to be played by outside stakeholders. One example of this is through the development of delivery chains, whereby all stakeholders are mapped to ensure a sound understanding of the inter-dependencies on the path towards successful outcomes. This resembles Elmore's backward mapping exercise, a key contribution of the second generation, bottom up scholars.

In terms of theoretical contributions by the policy implementation scholars, two decades of theory building have rendered some sense of the critical success factors. The Deliverology success variables coincide with the three categories of variables studied by second generation scholars: i) policy form and content, ii) organisations and the resources available as well as iii) the people responsible for implementation. In terms of policy form and content: ensure a few strategic priorities are selected that are outcome focused and supported across government. On the people front: strong, visible political leadership is imperative, as well as the establishment of a delivery unit that "has the ear of the political leader". In terms of organisational setting and resources: routines, backed by real-time data must be put in place. It is also important that a delivery unit understand the delivery system to ensure relevance to their work.

Deliverology as an evolving framework

The Deliverology framework is dynamic, with modifications and refinements made not only by the author of the approach (Michael Barber in conjunction with others), but also the delivery units that implement Deliverology. The first ever delivery unit – the PMDU – had five responsibilities, which encompassed broader policy work on the PSAs. In a 2011 publication, Barber in conjunction with McKinsey colleagues, present Deliverology as a six-step process, premised on the performance management cycle. In 2016, a more descriptive five step process with sub steps, targeted at educational leaders was presented by Barber and colleagues from the Education Delivery Institute. It is this recent framework that is utilised as the analytical framework for part two of this study.

It is also not uncommon for those instituting Deliverology to adjust the approach to fit their country contexts. The modifications can be categorised in two ways: a) adjusting the way in which a Deliverology step is undertaken or b) amending the functions undertaken by a delivery unit. In terms of the latter, the literature overview as well as the September 2018 study visit by the DSU, showed that delivery units tend to keep the basic steps (and functions) intact, with some additional functions added, rather than functions being omitted. Having said that there is still the risk that delivery units become "Delivery units in name only". This has led to the identification of the critical success variables that delivery units need to adhere to so they can be effective.

Despite the customisation and modification to Deliverology, the most commonly performed functions include i) assist with the articulation of priorities to be tracked, ii) develop key metrics and targets, iii) undertake data analysis and visualise progress in a meaningful way, iv) put in place reporting routines and v) utilise different tools to solve delivery problems.

The value add of the Deliverology approach

Deliverology is credited for turning vague electoral promises into reality and ensuring delivery receives the high-level attention it deserves. The authors of this approach have also developed many simple, user friendly tools that gives guidance during the planning, implementation, and performance tracking phases. In the main, the criticisms levelled against Deliverology are data related, which is not surprising given the strong emphasis on the performance aspects. This includes concerns surrounding the accuracy of data, the way in which targets are set, the selection of only a few key metrics, the challenges related to outcome measurement such as ensuring the indicators align to outcomes (validity) and inability to demonstrate causality (what was the delivery unit's contribution in achieving the desired results?). It is evident that the challenges surrounding outcome measurement, which features prominently in the performance measurement historical overview, are also found in Deliverology.

The Western Cape Government application of Deliverology

The DSU, like other delivery units, tailor-made Deliverology to make it accessible and understandable to those who would implement it. In adapting and modifying the approach, the DSU developed the 4P model. The 4P model presents the key principles of Deliverology: focus on a few strategic **priorities** of the government, have the right **people** involved, collect data and track **performance** and rapidly **problem-solve**. The 4P approach fully aligns with the common functions performed by delivery units: 1) ensure there is a clear articulation of priorities (giving strategic clarity to the priority choices of government); 2) develop the metrics and targets as well undertake data analysis and visualisation of data (performance principle); 3) establish the routine of meetings that are explicitly focused on addressing delivery blockages and problem solving; and, 4) ensure an engaged leadership that understands what is required of it and well-trained delivery teams with the necessary diversity of skills and expertise.

Our focus now turns to the **performance** principle of the DSU 4P model. The Western Cape context necessitated a more hands-on role in performance measurement, as will be demonstrated through the eLearning Game Changer. Four reasons necessitated this expanded emphasis on performance measurement in the DSU: a) officials' limited understanding of outcomes and how to measure outcomes, b) the limited availability of existing data to track the Game Changer performance, c) the lack of departmental capacity and resources to collect and analyse the data and d) the nature of priorities covered by the DSU, which were in most instances cross-cutting across departments and complex. This

led to a shift in the way data was approached: instead of the DSU fulfilling a support role, the DSU had to implement and manage many of the performance measurement processes.

I commence with a brief introduction of Part two before describing the selected case of this study - the eLearning Game Changer.

PART TWO

INTRODUCTION TO PART TWO: Towards an analytical framework

In the introductory chapter of the thesis, the main aim of my study was formulated as follows: To demonstrate that a modified Deliverology approach is an effective analytical framework to assess the performance of complex social interventions – viz. Game Changers. In order to address this research aim, the study was unpacked into two subsidiary research objectives: The first question (to describe the key features of the Deliverology approach as it has evolved elsewhere in the world and also how it was implemented in the Western Cape DSU) was extensively discussed in Part One of this thesis.

Part Two is devoted to a discussion of the second and arguably more important research objective of the study: to describe and reflect on how Deliverology was modified in the Western Cape and how these modifications – which were informed by my reading of key elements from mainstream programme evaluation traditions – produced clear and demonstrable gains in assessing the performance of the eLearning Game Changer.

In Part Two I discuss in each chapter the gaps and shortcomings of the existing version of Deliverology, followed by an extensive discussion of the main changes and modifications that were made in the course of the implementation and monitoring of the eLearning game changer. This discussion is organized around the steps typically included in Deliverology but with a focus on those steps which contained identifiable performance monitoring and measurement elements (and which was my responsibility as the Performance Tracking manager).

Figure 14 summarises the steps that are included in this study. The **BLACK bold font** denotes the main changes and modifications to the performance monitoring and measurement related steps of Deliverology as will be elaborated on in part two. The steps in **RED** font were undertaken jointly by the DSU and the two implementing departments (WCED and CeI) but will not be covered by this study.

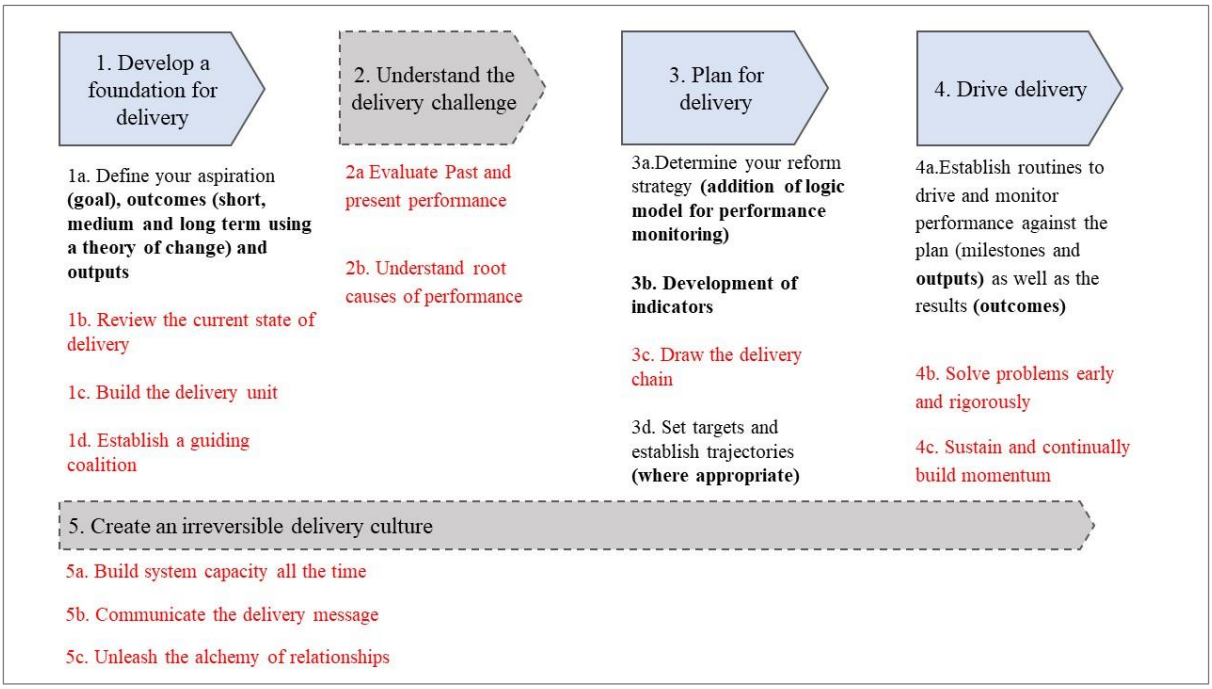


Figure 14: Summary of modifications to the Deliverology framework as well as steps excluded from this study

The reasons for excluding certain steps are discussed below (**RED font** as per Figure 14):

Step 1: Develop a foundation for delivery

- Step 1b entails an assessment of the current state on the delivery. Utilising the steps, and sub steps as the framework each sub step was colour coded using a four-tiered assessment framework to reflect the status of that sub step at a certain point in time. The DSU undertook this assessment quite frequently, especially during the planning phases when data on Game Changer progress was not yet available. The assessment framework provided a snapshot of how the Game Changer was progressing in executing each of the five steps, i.e. "building the delivery unit" was initially scored amber red, as it took time to put the full staff complement in place. This sub step addresses the progress in getting the Game Changers "off the ground". This includes checking progress in putting in place the performance measurement related steps (i.e. setting targets and trajectories, developing the delivery plan etc.), but does not cover the detail of how the performance measurement was executed. For this reason, this step is excluded from this study.
- Step 1c deals with building a delivery unit as this is the mechanism by which Deliverology is institutionalised. I have already discussed this earlier in the chapter and specifically how the

DSU translated Deliverology into a 4 "P" model, which includes prioritisation, people, performance and problem solving.

- Step 1d is related to building a guiding coalition. Although it is not covered in this study, a guiding coalition does strengthen the foundation for delivery. Barber (2015, p.54) describes this as follows "the guiding coalition is not the same as a management team; it's a shared understanding among seven to ten people in key positions about what needs to be done and how". A relationship is established with a core group of individuals where honest conversations around implementation can take place. It presents a team of people "you can count on" that shares "a collective commitment to the aspiration" (Barber, et al., 2016, p.57). Building these relationships were part and parcel of the DSU's work – in particular the DSU lead (Ms Jenny Cargill) and the eLearning Game Changer lead (Ms Penny Tainton). The two monthly stocktakes and preceding MANCO meetings were formal meetings and served a different purpose than these "coalition building" meetings. Informal meetings were convened often to discuss some of the "sticky issues" directly with the political and administrative head which assisted in unblocking some persistent delivery challenges.

Step 2: Understand the delivery challenge

- Step 2 deals with the need to understand the delivery challenge recommending that past and present performance be evaluated (sub step 2a), as well as that the root causes of performance be interrogated (sub step 2b). In our selected case, the eLearning Game Changer, this was not done as a dedicated step given that the implementation of e-Education was already underway. However, past and present performance was considered during step 1, when the goal statement for the eLearning Game Changer was formulated. Aside from this, I do not cover this step in any substantive detail.

Step 3: Plan for delivery

- Step 3b refers to the inclusion of delivery chains as part of the delivery plan. A delivery chain is defined as "the set of actors (people or organisations), and the relationships between them, through which a given strategy in your delivery plan will be implemented" (Barber et al., 2016, p.143). A delivery chain exercise is usually done for each strategy, starting with the end goal in mind and working one's way back to plot the various actors and their role in ensuring the achievement of the ultimate goal. Delivery chains serve multiple purposes: when a breakdown in delivery occurs the delivery chain can be utilised to identify the person responsible. In addition, delivery chains assist in creating a shared understanding of how the strategy must be

delivered as well as developing milestones. Delivery chains were developed for the eLearning Game Changer but does not have direct relevance to this study. Hence, this sub step was not included in this study.

Step 4: Drive delivery

- Step 4b: problem solving is part and parcel of the Deliverology approach, with stocktake routines and data being utilised to identify the problem and to act quickly. This was also the case with the eLearning Game Changer. However, I did not modify this step in any way and hence it does not form part of this study.
- Step 4c: implementation is characterised by up and down cycles, making it important to ensure momentum is sustained. Some useful recommendations are provided in this regard, but this step links directly to the delivery of the eLearning Game Changer, and not the performance measurement aspects. It was therefore excluded from this study.

Step 5: Create an irreversible culture (crosscutting step)

- Step 5a is about building the system's capacity so government do not return to "business as usual" once delivery units are terminated. Although the DSU instituted this to some extent by means of capacitating teams on the Deliverology approach, convening hand over meetings and addressing aspects such as human resources and project budgets beyond the Game Changer period, ultimately the new Premier opted not to continue with the Deliverology approach (May 2019). One must recognise the time limitations of creating irreversibility: instilling a new culture and changing engrained work habits would take significantly longer than the three and a half years the Western Cape DSU was in existence. The performance measurement aspects were included in the hand over process. However, it does not have relevance to this study and is therefore not included.
- Step 5b: The communication of the DSU was focused on campaigns in support of the Game Changer programmes, not internal government communication. Even though greater emphasis on the internal communication would have probably benefitted the Game Changers, a substantive new culture of delivery was ultimately required. The DSU identified the lack of change management in the Deliverology approach as a gap. The communication and change management aspects have no bearing on the topic of this study and is therefore excluded.
- Sep 5c is about relationship building. As Barber et al (106, p. 338) state "relationships are the glue that will hold your delivery effort together". Relationship building was part and parcel of the DSU's work. The building of relationships with district staff and sample schools in support

of the performance measurement requirements is dealt with in Chapter 8 and does not warrant a separate discussion.

Given that part two focuses on the modifications to the Deliverology approach, I suffice with a summary overview of the modifications and additions (**black bold font** as per Figure 14):

Step 1: Develop a foundation for delivery

- Step 1a: in "defining our aspiration" I incorporated key elements of a theory-based approached to evaluation. More specifically I added to this step the essential need for the project team to develop a theory of change for the eLearning Game Changer which articulates a) the causal pathway(s) of the programme and b) also introduces a clear distinction between short, medium and long term outcomes. A detailed discussion of these changes is found in Chapter 6.

Step 3: Plan for delivery

- Step 3a deals with the development of reform strategies as part of the delivery plan. Deliverology is not prescriptive about the format of the delivery plan but does make suggestions around the elements to be included, i.e. reform strategies, targets and trajectories and delivery chains (an excluded sub step). With Step 3a I utilised some of the core terminology of the 'logic model framework' to develop a coherent delivery plan (which we refer to as a roadmap) for the eLearning Game Changer. Chapter 7 documents these changes.
- I introduced a new step (Step 3b) to capture the development of indicators: the selection of indicators (or metrics as Deliverology refers to it), are discussed as part of other steps in the Deliverology framework. Some guidance is provided on how to select indicators, as well as the different categories of indicators to consider. However, Deliverology largely assumes the use of existing indicators and therefore does not cover the conceptualisation or development of (new) indicators in any detail. Bearing in mind the challenges associated with outcome measurement, I therefore added an additional step to Deliverology (new Step 3b) that draws on the programme evaluation paradigm as well as the lessons learnt from, for example, the social indicator movement. This additional step provides a detailed overview of how indicators for the eLearning Game Changer were i) conceptualised and ii) operationalised. Chapter 8 is devoted to this discussion.

- Step 3d: the "original" step 3c (now 3d) requires for targets and trajectories to be developed for all the "goals". We did not set targets in all instances for the eLearning Game Changer. Given that many of the outcomes were being tracked for the first time no baseline was available. A decision was taken to utilise the Game Changer period to establish the baseline against which future target setting could be done. I discuss the use of targets in the eLearning Game Changer in Chapter 8.

Step 4: Drive delivery

- As far as Step 4a (Establish routines to track and monitor performance) is concerned I drew on the performance measurement tradition to provide more substance to the process of performance monitoring. Although the Deliverology approach requires that progress in implementing the delivery plan and the achievement of results be tracked on a continuous basis, it does not distinguish clearly between 'output' and 'outcome' monitoring. I argue that this distinction is important – not only does it align with a theory-based approach to programme evaluation but also with the distinction between short- and medium-term outcomes in our outcome mapping. Chapter 9 provides an overview of this discussion, drawing on the eLearning Game Changer output and outcome data.

The subsidiary research questions align to the Deliverology framework, as well as the chapter breakdown discussed above. The table below provides a summary:

Table 9: Subsidiary research questions for Part two

RESEARCH QUESTION 2: How was the Deliverology approach modified in the Western Cape Government, and what gains did these modifications "produce" in assessing the performance of the eLearning Game Changer?	
Step 1a: How does the inclusion of a clarificatory evaluation step (by way of a Theory of Change) contribute to a better understanding and monitoring of the eLearning Game Changer?	**Chapter 6**
Step 3a: How does the inclusion of an explicit logic model improve the monitoring of the eLearning Game Changer?	**Chapter 7**
Step 3b: How does the introduction of a clear distinction between outcomes and performance indicators enhance the monitoring of the eLearning Game Changer?	**Chapter 8**
Step 4a: How does a clear distinction between performance monitoring and outcome monitoring assist in reporting on the findings of the eLearning GC?	**Chapter 9**

My role as participant observer

The ability to undertake the modifications and additions to the Deliverology framework was made possible through my role as a participant observer. Participant observation is typically described as a qualitative data collection method with roots in anthropology, in particular ethnographic studies (Kawulich, 2005). Schensul, Schensul and LeCompte (1999, as cited in Kawulich, 2005, p. 2) define participant observation as "the process of learning through exposure to or involvement in the "day to day" or routine activities of participants in the researcher setting". Although I was not studying people and collecting data in the manner commonly associated with participant observation, my close involvement in instituting the Deliverology approach, and constant contact with the implementing teams allowed for an incremental learning and deeper insight as the process unfolded. Through this 'participant observation" process I was able to identify and repeatedly validate the potential shortcomings. In this process I also drew on the knowledge and understanding I gained in my programme evaluation work.

One of the main challenges associated with the participant observation approach is that of maintaining a professional distance, while immersing oneself in the phenomenon being studied (Takyi, 2015). I navigated this challenge by taking on a "participant as observer" role, which Takyi (2015) views as the preferred role compared to the other three roles of the participant observer[8]. The "Participant as observer" role is characterised by the development of relationships with the informants of the study, as well as being intricately involved in the "insiders' central activities (Baker, 2006, as cited in Takyi, 2015, p. 868). In addition, the participant as observer tends to make his/her position as researcher known which removes suspicion and builds rapport between the researcher and the informants.

This is an appropriate description of my role as the performance tracking director in the DSU. My position was known to the implementing teams and being responsible for all performance tracking activities over a three year period (including the clarificatory elements) meant relationships with the implementing departments were established. In my role as "participant as observer role" I acquired context specific knowledge of what would work and what would not from a performance measurement

[8] Gold (1958) identified four roles of a participant observer: the complete observer, the complete participant, the observer-as-participant and the participant-as observer. The complete observer only observes, i.e. he/she does not participate with the research subjects. The complete participant researcher on the other hand, fully participates and therefore tends to not reveal his/her identity as a researcher. The observer- as-participant tends to observe more, maintaining a professional distance throughout.

perspective. This, together with my programme evaluation background, shaped the way in which I instituted the performance measurement elements of the Deliverology – ultimately modifying and introducing some additions to the Deliverology framework.

Before I turn to a discussion of each of the subsidiary research questions, included under the second objective, it is necessary to provide a description of the eLearning Game Changer selected as the illustrative case for this study. The next chapter discusses the design and methodological elements of the eLearning Game Changer.

Chapter 5: The eLearning Game Changer: an illustrative case study

5.1 Introduction to the case study design

As indicated in the first chapter of the thesis, the eLearning Game Changer was selected as the object for this study. Like many educational interventions, the eLearning intervention can be classified as a complex social intervention. The distinction between simple, complicated and complex programmes was originally introduced by Patricia Rogers in a highly cited paper (2008). In a recent paper, Chatterji (2016) presents a more elaborate and systematic discussion of what complex interventions are. He summarises the main elements as follows (2016, p. 131):

> Complex social programmes are socially mediated service interventions. That is they are delivered by human agents, either individuals or socially-organised groups, with various levels of organisation and autonomy in the processes they employ for service delivery. Complex social interventions are also typically multivariate in configuration with many moving parts and pieces. Generally, complex social programmes operate within larger, multi-level community structures or organizational systems with nesting of social units, and different degrees of openness in the flow of activities and people. Multiple actors and agendas could be at play that directly or indirectly influence the functioning of the complex social programme as well as the observed outcomes.

As will become clear in our discussion of the content of the eLearning intervention, it manifests all the features of complex interventions. Against this background, a case study design was deemed to be the most appropriate type of research design for my study. Case studies are a popular design option in the social sciences, specifically in "practice-oriented fields" such as education (Starman, 2013, p. 29). In addition, case studies are particularly appropriate design types for capturing the complexity of highly contextualised phenomena (Rosenberg & Yates, 2007; Yin, 2003). The definition by Simons (2009, cited in Thomas, 2011, p.512) reflects this well:

> Case study is an in-depth exploration from multiple perspectives of the complexity and uniqueness of a particular project, policy, institution, program or system in a "real life" context.

Despite a substantial body of scholarship on the topic, "case study design" still means different things to different people, resulting in what Gerring (2004, p. 342) refers to as a "definitional morass". One explanation offered for the challenges associated with defining case studies, is the different

epistemological starting points of scholars utilising case studies. Case studies have a long history, with the rise of case study research associated with anthropology and the social sciences (Johansson, 2003). The Chicago School of Sociology is recognised for their ground-breaking work in case study research between the 1920s and 1950s whereby they undertook extensive field-based observations (Stewart, 2014 cited in Harrison, Birks, Franklin & Mills, 2017). Viewed as qualitative research early case studies aimed to document people's experience and sense-making within their natural settings (Merriam, 2009 cited in Harrison, Birks, Franklin & Mills, 2017). With the onset of positivism in the 1950s, quantitative methods became the favoured approach, with case studies either forming part of the quantitative study or being utilised in a narrow way to describe a certain phenomenon (Johansson, 2003). The rise of grounded theory through the work of Glaser & Strauss (1967) paved the way for a new generation of case study research (Harrison et al, 2017). Grounded theory combines qualitative fieldwork methods with quantitative data collection methods, which led to an inductive approach to case study research. In building on this work, Yin applied experimental logic which allows for theory testing and causal analysis to be done within case study research (Harrison et al, 2017).

The recognition that case studies represent a design choice as opposed to a methodological choice were not always self-evident (Rosenberg & Yates, Thomas, 2011). In more recent literature there appears to be consensus that the methodological choices - whether quantitative or qualitative or mixed – are a secondary concern to a decision on design type as demonstrated by Stake's description of a case study:

> [A] Case study is not a methodological choice but a choice of what is to be studied… By whatever methods we choose to the case. We could study it analytically or holistically, entirely by repeated measures or hermeneutically, organically or culturally, and by mixed methods – but we concentrate at least for the time being, on the case (Stake, 2005, p.443 cited in Thomas, 2011, p.512).

The case study design offers many advantages, which include greater conceptual validity, the ability to develop a new hypothesis and explore causal mechanisms in greater detail compared to other design options. The conceptual validity speaks to the advantage offered by case studies to pay closer attention to the contextual elements when identifying, and tracking "hard to measure" social phenomena. Even though data collection might not be as extensive as in quantitative research design studies, the results from the case study could have higher validity because of "contextualised comparison" being performed, "which automatically searches for analytically equivalent phenomena even if they are expressed in different terms and contexts" (Starman, 2013, p.36).

It is also necessary to debunk some common myths associated with case study design: firstly, that contextualised case study learnings are sub-standard compared to the production of general, theoretical knowledge, and secondly that case study research has limited generalisation potential. In addressing the first myth, it must be recognised at the outset that the social sciences have had limited success as far as general theory building is concerned (Starman, 2013). Of much more use is context-specific practical knowledge as captured through case studies (Starman, 2013). In terms of the generalisation argument, case study designs seek analytical generalisation as opposed to statistical generalisation. The latter is associated with large scale survey studies whereby empirical data is collected from a sample of the population with the aim of generalising the findings to the bigger population. In case study design the "sample" of cases is not selected with the aim of statistical generalisation, i.e. concluding on the basis of the selected case's findings that the same results would apply to the broader population. Rather, in case study design the purpose is to demonstrate analytical generalisation, provided "the salient features of the case are documented so that new situations can be illuminated by a very thorough understanding of a known case (Sturman, 1997, cited in Starman, 2013, p. 39).

5.1.1 Thomas' classification of case study designs

Various scholars have attempted to organise and categorise different types of case study design. In summarising some of the renowned scholars' case study typologies (Table 10 below), Thomas (2011) notes that these typologies tend to combine different criteria and layers. For example, one of the criteria emerging through the typologies is the distinction between theoretical and non-theoretical case studies (see for example George and Bennet, Bassey). This distinction represents a different layer than for example distinctions made between methodological choices (i.e. Yin who refers to longitudinal case studies).

Table 10: Case study typologies

George and Bennett (2005)	Merriam (1998)	Stake (1995)	Bassey (1999)	De Vaus (2001)	Mitchell (2006)	Yin (2009)
Theory testing	Descriptive	Intrinsic	Theory seeking	Descriptive/ explanatory	Illustrative	Critical
Atheoretical/ configurative-idiographic	Interpretative	Instrumental	Theory testing	Theory testing/ theory building	Social analytic	Extreme/ unique
Disciplines configurative	Evaluative	Single/ Collective	Storytelling	Single/ multiple case	Extended	Longitudinal
Heuristic			Picture drawing	Holistic/ embedded	Configurative-idiographic	Representative

George and Bennett (2005)	Merriam (1998)	Stake (1995)	Bassey (1999)	De Vaus (2001)	Mitchell (2006)	Yin (2009)
Plausibility probes			Evaluative	Parallel/ sequential	Disciplines configurative	Revelatory
"Building block studies"				Retrospective/ prospective	Heuristic	
					Plausibility probes	

(Thomas, 2011, p. 516)

In response to the different typologies, Thomas (2011) presents his own typology: one which I have adopted as the point of departure for the classification of my case study. The figure below presents Thomas' framework.

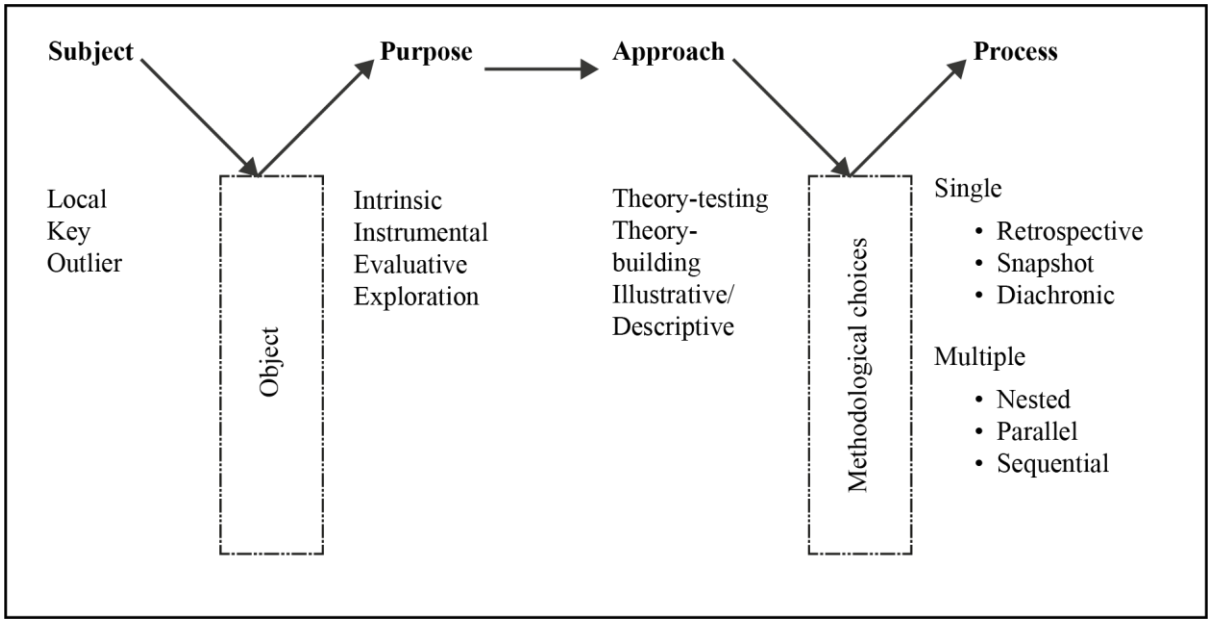


Figure 15: Thomas' typology of case study research

(Thomas, 2011, p. 518)

In his framework Thomas (2011, p. 513) begin with a strong distinction between the subject and the object of the case study. He suggests that a case study must comprise two elements:

1. A "practical, historical unity," which I shall call the *subject* of the case study, and
2. An analytical or theoretical frame, which I shall call the *object* of the study.

Taking this distinction into account, his first version of a definition of a case study is as follows:

> Case studies are analyses of persons, events, decisions, periods, projects, policies, institutions, or other systems that are studied holistically by one or more methods. The case that is the subject of the inquiry will be an instance of a class of phenomena that provides an analytical frame—an object—within which the study is conducted and which the case illuminates and explicates (2011, ibid).

He continues to elaborate on this distinction: "Whereas the subject can be quite broad, the object must be quite specific to indicate "what is this a case of". The subject can be any phenomenon, however broad or narrow whereas the object needs to provide the analytical frame for the case study. The link between the subject and object is as follow: "in order for it to constitute research there has to be something to be explained (an object) and something to offer explanations (the analysis of the circumstances of a subject)" (Thomas, 2011, p.513).

According to Thomas, there are three potential routes for selection of the subject. The first route in its selection may be followed because of the researcher's familiarity with it—a *local knowledge case*—and this will be relevant particularly for the practitioner or student researcher. The second route is followed when the case is interpreted as a key or illustrative case; and the third route is followed when the case is deemed to be an outlier case. The key case is selected because of its ability to highlight very particular issues surrounding a particular phenomenon, whereas the outlier case would offer the opportunity to capture the deviance or differences from a common pathway or commonly accepted practice (Thomas, 2011).

Drawing on the typologies presented in Table 10, Thomas (2011) extracts four purposes off case studies: intrinsic, instrumental, evaluative and exploration. The intrinsic and instrumental purposes are borrowed from Stake (2005). Intrinsic case studies are not focused on theory building but are selected because "the case itself is of interest" (Baxter & Jack, 2008, p. 548). With instrumental cases, the case itself is not the primary concern but is used to better understand something else (Baxter & Jack, 2008). Exploratory case studies are undertaken to "enlighten those situations in which the intervention being evaluated has no clear single set of outcomes" (Yin, 2009, p.20). Thomas does not propose a specific definition of 'evaluative' but simply refer to the work of Merriam (1988).

The next phase or dimension in Thomas' typology represents the approach to the study. The case study can either be approached from a theory testing, theory building or descriptive perspective. This

categorisation captures the distinction between theoretical and nontheoretical case studies, with descriptive case studies falling within the nontheoretical category (Thomas, 2011). Once the approach has been clarified Thomas argues that the researcher now needs to make specific methodological choices. This includes decisions around case selection (units of observation), which measurement instruments to use, how to collect the data and how to analyse the data collected.

This brings us to the final element namely process, which covers the operational processes of the case. It is necessary at this point to revisit the subject, as well as any case parameters that were established at the outset to bound the case. This entails consideration of the time period, place, event, institution, person or any phenomena. The biggest consideration is whether the process of studying the subject constitutes a single or multiple case study. A multiple case study implies that a comparison of different cases will be undertaken, which is further sub-divided by Thomas into nested, parallel or sequential case studies (Thomas, 2011). Single studies take on three forms: retrospective (after the fact), snapshot and diachronic. The time dimension is evident in all three single studies: retrospective studies cover past occurrences or phenomena, snapshot studies aim to capture the subject at a specific point in time, while diachronic studies capture change over time (Thomas 2011).

5.1.2 Application of Thomas Typology to the eLearning Game Changer

In this section, I indicate how I have applied Thomas' framework for classifying and depicting a specific kind of case study. The application of Thomas' typology to the eLearning Game Changer case study design is indicated by the bold-bordered text boxes in Figure 16:

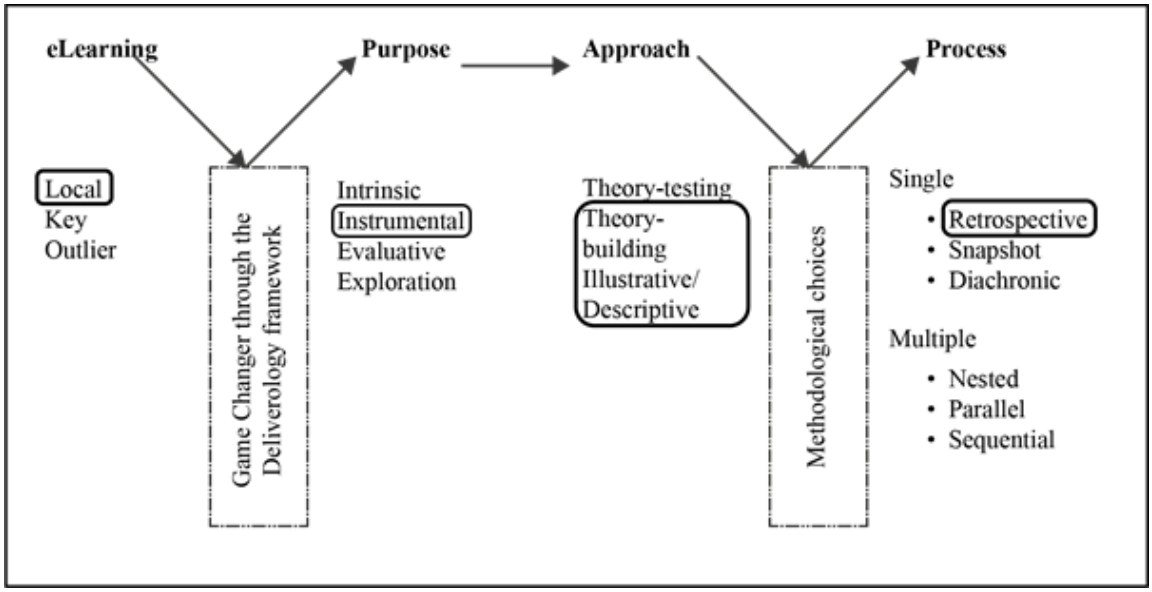


Figure 16: Thomas' typology as applied to the eLearning Game Changer

I discuss each element of Thomas' typology as applied to my study in more detail.

Subject

As already indicated in the first chapter, the subject of my case study is eLearning (in Thomas' framework the phenomenon or domain under investigation), whereas the object of our analysis is the eLearning Game Changer as conceptualised and defined within the analytical framework of Deliverology. In this section, I give some more background and context to why I chose the field of eLearning as the subject of our case study.

The start of eLearning in the national education sphere in South Africa can be traced back to the mid-Nineties when the Technology Enhanced Learning Initiative (TELI) was launched in 1995. This was followed by the development of the TELI strategic plan (1997) which identified six projects in support of a technology enhanced educational environment (Howie, 2010).

A broader national response to ICT was strengthened by the establishment of the Presidential National Commission on Information Society and Development as well as a Presidential Advisory Council on Information Society and Development (both in 2001), consisting of experts in the field with the role of advising government on the roll out of ICT (DOE, 2004). The national government department responsible for all ICT initiatives, the Department of Communications, identified the need for an e-

Strategy to be developed that would include the education sector (DOE, 2004). Work on a national ICT strategy commenced in 2001 by means of a National ICT forum (Howie, 2010) and in 2013 the e-Education strategy was released.

The 2004 eEducation White Paper on ICT acknowledges the need for education to keep track of global ICT developments and to ensure learners and schools have access to the multiple opportunities offered by ICT (DOE, 2004). These opportunities include the potential to increase administrative efficiency, allow better communication and collaboration, act as a resource for curriculum integration and enable a more creative and engaging learning environment (DOE, 2004).

The first major eLearning project in the Western Cape was the Khanya project. This project was implemented in 2001 by the WCED with the aim of "making technology accessible to all learners in the Western Cape" (du Toit, 2005, p.1).

In 2012, the WCED launched their eVision which had a two-fold focus: to enhance teaching and learning and to produce learners that are prepared for the 21st century's world of work (WCED, 2015, WCED, 2012). The eVision leveraged off the provincial government's broadband programme which was started in 2009 and which aimed to provide high speed broadband access to 1 900 government sites (that include schools).

The provincial e-Education plan provided the work plan for the eVision, identifying 14 projects to be implemented in a phased approach over 5-year increments. The reason for this is that budgetary and resource constraints prevented a province wide implementation. The e-Education plan covers the 2015/16 to 2019/2020 financial years and provides annual targets and budget requirements for all 14 projects. Implementation of e-Education was therefore well underway before the Western Cape DSU Game Changer programmes were launched.

The Western Cape Premier's keen interest in the education sector, as well as growing concerns around the utilisation of the WCG broadband programme, led to a decision to include eLearning as one of six project priorities under the purview of the DSU.

Object

It is not always clear what Thomas means when he refers to the 'object' of a case study. In one sentence he defines the object as "an analytical or theoretical frame, which I shall call the *object* of the study" (Thomas, 2011, p…515), and in another he writes "the subject can be any phenomenon, however broad or narrow whereas the object needs to provide the analytical frame for the case study" (Thomas, 2011, p 515). These two statements are seemingly at odds with each other: the first equates the 'object' of the study with the analytical frame or lens through which the subject is studied; the second, states that the 'object' provides the analytical frame for the case study.

Rather than getting stuck to clarify this issue, I decided on the following definition of our object of study: the eLearning Game Changer as conceptualised within (the analytical framework) of Deliverology. I, therefore, tend to veer towards his definition of the 'object' as the framework through which the subject is studied – in our case the Deliverology approach. In the remainder of his article, Thomas seems to prefer this interpretation when he argues that the object need not be defined at the outset but, rather, may emerge as an inquiry progresses.

> Whether it is set at the outset or is emergent, it will be this analytical focus that crystallizes, thickens, or develops as the study proceeds: It is the way that this "object" develops that is at the heart of the study. Whichever -"emergent" or set at the outset - it is important to have some notion of a potential object in mind when the study begins and not to confuse it with the subject (Thomas, 2011, p.514).

It is certainly the case that our 'object' was to a large extent set at the outset when the eLearning intervention (together with the other interventions initiated in 2015) was defined as a **Game Changer, to be implemented within the framework of the Deliverology approach as described in the introductory section of Part two**.

Purpose

In terms of Thomas' framework, our eLearning case study is primarily instrumental in purpose. The eLearning Game Changer is not the main concern; rather it is used as an example to document the results of applying the modified Deliverology approach to the Game Changer programmes. In the context of this study my purpose is specifically more formative (than summative) in nature: improving and guiding the implementation of the eLearning Game Changer.

Approach

The approach followed in this case study includes two of the categories in Thomas' framework. On the one hand, the case study can be categorised as illustrative/ descriptive in nature in that I document in detail the process and results of the performance monitoring undertaken in the eLearning Game Changer. The methodological choices in executing the eLearning Game Changer performance measurement aspects are expanded on in section 5.2.

On the other hand, I would also argue that it can be categorized as an 'exercise' in theory building, given our main argument that the Deliverology approach can be strengthened through the introduction of programme evaluation and performance measurement elements (for example clarificative evaluation and performance monitoring).

Process

In terms of Thomas' final criterion, the selected case can be described as a single, retrospective case study. It is categorised as single case, as no comparison with another case is undertaken. The case study is retrospective given that it is written up after the eLearning Game Changer work under the DSU had been concluded.

5.2 Methodological choices to measure the performance of the eLearning Game Changer

In this section I elaborate on the key methodological choices that were made to measure the performance of the eLearning Game Changer. I discuss the various methodological choices under four headings: i) selecting the cases to be observed within the study, ii) deciding what measurement instruments to use, iii) deciding how to gather the required data and iv) deciding how the data analysis would be done.

5.2.1. Case selection

The first methodological choice concerned the selection of schools to be included in the eLearning Game Changer as data could not be collected from all 1 500 public schools in the province. A major challenge at the time of launching the Game Changers was the severely constrained fiscal situation in the Western Cape Province. In the case of the eLearning Game Changer, it meant that the budget was not available to roll out the same level of technology to all of the approximately 1 500 public schools. On advisement from Sir Michael Barber, a categorisation of schools was done to reflect different levels

of technology roll out. This allowed for a phased approach to technology roll out, which took into consideration budget availability. The ±1 500 public schools were hence assigned to three categories based on the intensity or availability of technology platforms to support eLearning:

- Universal Schools (approximately 1 200 schools): which already had basic access to the internet through a computer laboratory or an ICT suite. In these schools, the WCG provided teachers and learners with access to digital resources through the Wide Area Network (WAN) via the computer laboratories and central Wi-Fi access points
- Enhanced Schools (approximately 370 schools): schools that already had a Local Area Network (LAN) implemented at that stage or which were earmarked to receive a LAN fell within the Enhanced category. The LAN allowed for every instruction room in the school to be connected to the WAN and therefore, allowed wireless access in the classrooms. In order to maximise on this access, teacher technology was prioritised in these schools.
- Model Schools: a handful of schools (16) which were provided with the full suite of wireless and other technology to cover the entire schools, with all learners receiving a device. The intent was to utilise the experience of these schools to support and drive eLearning throughout the province.

In selecting the cases to be "observed", a dual approach was followed: where new data was needed, both purposeful and probabilistic sampling was undertaken. Where existing data was available for all schools, the full data set was utilised (i.e. no selection or sampling was undertaken).

The sampling was undertaken within the broader Game Changer context. As indicated above, the DSU had oversight over three educational linked Game Changers: eLearning Game Changer, the After School Game Changer and the Apprenticeship Game Changer (referred to collectively as the Youth Game Changers). All three required data to be collected from schools. It was important not to overburden schools with data requests and therefore a 'pragmatic' approach to the sampling was undertaken. One hundred and eleven schools were subsequently selected to ensure data could be collected for the Youth Game Changers. In terms of the eLearning Game Changer, 63 schools were selected (out of the 111) from which to collect eLearning specific data. A total of 30 schools covered the After School Game Changer data needs and 20 schools covered the Apprenticeship Game Changer. The reason for these not totalling up to a 111 is that there was some overlap of schools across the three Game Changers. The figure below summarises the sampling plan for the three youth Game Changers, also providing the breakdown of 63 eLearning sample schools across model, enhanced and universal schools.

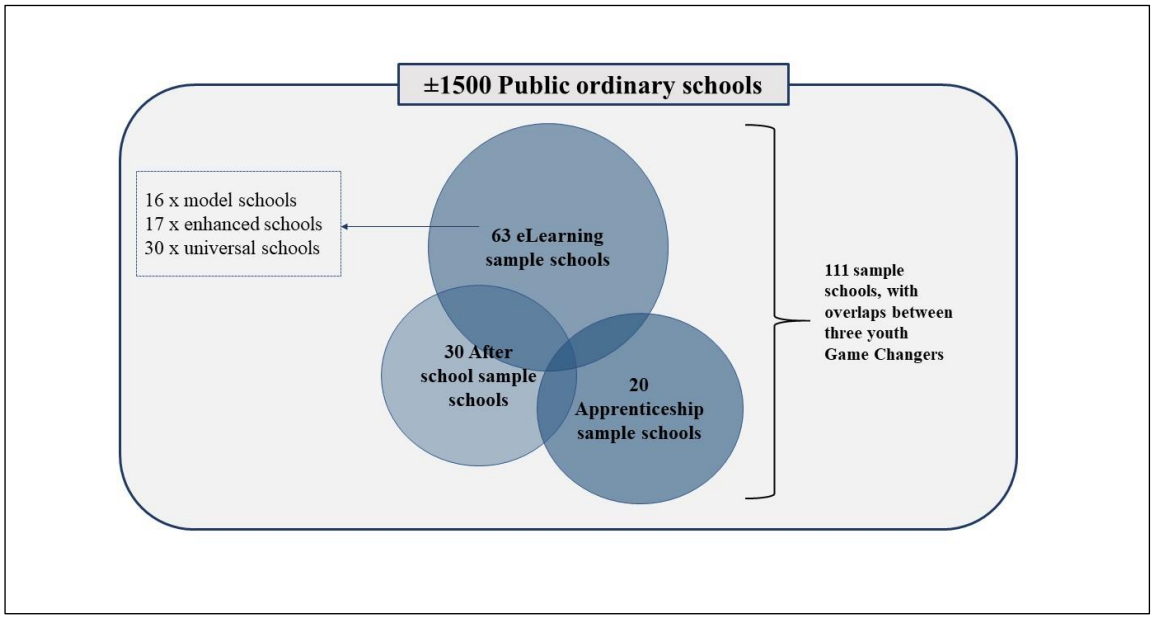


Figure 17: Summary of sampling of three youth Game Changers, including eLearning

In selecting the 63 eLearning sample schools, stratified multi stage sampling was undertaken. The first stage entailed sampling a suitable number of schools. I utilised two variables to arrive at the number of 63 eLearning sample schools: school category (model, enhanced and universal) as well as educational district. The Western Cape has eight educational districts, and it was important to ensure a spread of schools across the eight districts to include rural and urban schools. Table 11 provides the breakdown of sample schools across the districts, per school category.

Table 11: First stage in the sampling

School category	No of schools	Metro Central	Metro East	Metro South	Metro North	Cape Winelands	Overberg	Eden & Central Karoo	West Coast	Sampled sites
Model	Population	2	2	2	2	2	2	2	2	16
	Sample	2	2	2	2	2	2	2	2	**16**
Enhanced	Population	49	65	49	40	47	13	29	13	305
	Sample	2	3	2	2	2	2	2	2	**17**
Universal	Population	165	115	156	156	233	72	194	117	1208
	Sample	4	2	4	4	6	2	5	3	**30**

The second stage of sampling entailed selecting the actual schools per district that would be approached to participate in the data collection. For this: three further variables were taken into consideration:

- The quintile of the schools (which gives a broad indication of the socio-economic status of the learners in the school[9]);
- The performance of schools (using learner progression performance statistics[10] and Grade 12 examination result) to ensure the inclusion of both struggling and well performing schools;
- The commitment of the principal was also considered as it was important that the selected schools commit to the data collection over a two-year period. Some would argue that this could introduce bias as strong leadership is viewed as one of the critical success factors associated with well performing and successful schools. The team however argued that it was more important to have the schools commit to the data collection process for two years than run the risk of schools dropping out mid-way, thus compromising the data efforts entirely.

The intent was to increase the sample size every year but due to resource constraints this did not materialise.

5.2.2 Measurement

The measurement choices of this case study refer to the development and selection of suitable instruments to gather information about the selected cases (schools). A distinction is made between highly structured forms of measurement compared to semi and unstructured forms of measurement. The highly structured forms of measurement are associated with the classical meaning of measurement whereby structure is imposed on the observations beforehand through predefined categories which allows for easy quantification and analysis of the phenomenon. Semi to unstructured forms of measurement constitute the opposite side of the spectrum with little or no structure imposed beforehand and typically results in narrative or textual data. This, off course, does not limit the quantification possibilities, but with the difference being that the narrative data is coded and subsequently quantified after the observations have taken place.

[9] The quintile system allocated all government schools into one of five categories, with quintile 1 schools designating the poorest institutions while quintile 5 denoted the least poor public schools. The quintile to which a school was assigned was based on the rates of income, unemployment and illiteracy within the school's catchment area. Source : https://www.corruptionwatch.org.za/schools-quintile-system-to-change/. Accessed 17 September 2019

[10] Progression statistics show the number of learners that have moved to the next grade

In the eLearning Game Changer, only highly structured forms of measurement were undertaken. This consisted of a) scales and scaled items (i.e. predefined categories of the phenomena being measured) and b) a comprehensive list of indicators to track performance. I discuss these two forms of 'measurement' (qualitative and quantitative) in more detail in Chapter 8. An example from the eLearning Game Changer will demonstrate this: on the qualitative front, I was interested in determining teacher, learner and principals' views and perceptions of the eLearning roll out. From principals I wanted to establish their level of commitment to eLearning, but "commitment" cannot be directly observed. Through the process of operationalisation, the concept of commitment was unpacked, which led to a set of questions that were then included in a questionnaire. A significant part of the eLearning Game Changer consisted of tracking different types of indicators. Examples of this included tracking the roll out of learner devices, teachers being trained and schools receiving the WAN.

5.2.3 Data collection methods

At the highest level, data collection methods can be divided into two broad categories: non-reactive or unobtrusive methods of data collection and reactive measures of data collection. With non-reactive methods, the data collection is in no way obtrusive to the person or entity being studied. Examples of this include existing data and statistics. With reactive measures, the person or entity being studied is aware that they are being observed, which can in turn influence the quality of the data gathered. Reactive measures can be further subdivided into observations and self-reporting methods (for example interviews and surveys).

In the eLearning Game Changer both non-reactive and reactive data collection methods were utilised. Existing data as well as statistics were collected, and on the reactive side the intent was to conduct classroom observations as well as undertake surveys. All data-gathering methods were highly structured.

The table below overlays the sampling and selection choices with the data collection methods as well as data collection tools (distinguished between reactive and non-reactive). For some of the data collection methods, data was already being collected or an existing data collection platform was already in place. However, the data collected through these existing systems did not meet the Game Changer requirements which required that certain adjustments needed to be made. These data categories have been indicated as "customised data", given that changes needed to be made to fully accommodate the Game Changer requirements (for example the principal questionnaire which had to be modified to include eLearning specific questions).

Table 12: Data collection methods (and tools): split between reactive and non-reactive

Data collection methods and tools		New/ Existing/ Customised	Sampling/selection of schools (cases)
REACTIVE (OBTRUSIVE)	**Surveys**		
	Teacher and learner questionnaires	New data	63 schools
	Principal questionnaire	Customised data	All schools
	Timetable schedule	New data	47 schools
	Observations		
	Classroom observation schedule	New data	16 schools
NON-REACTIVE (UNOBTRUSIVE)	**Platform and implementation data**		
	Learner and teacher level WAN usage report	New data	111 schools
	WAN Downtime report	New data	111 schools
	eCulture checklist	New data	All schools
	Centralised educational management information system (CEMIS) report	Customised data	All schools
	Technology roll out checklists	Existing data	All schools
	CEI Service desk data	Customised data	All schools
	ePortal usage data	Customised data	All schools
	Learner and teacher data		
	Learner level school attendance and performance data	Existing data	111 schools
	Teacher training data	Existing data	All schools

Each of the data collection methods and its accompanying tools are described in more detail below:

Surveys

Questionnaires: teachers, learners and principals

The development of all new data collection tools was led by the performance tracking manager in the DSU with input obtained from both WCED and CeI. The content of the learner, teacher and principal questionnaires are summarised in the table below. The three questionnaires were all administered utilising online platforms.

Table 13: Content of the questionnaires

Learner questionnaire	Teacher questionnaire	Principal questionnaire
• Access to technology and internet at home • Access to technology at school • ICT proficiency • Access to digital content • Experience of digital content • Attitude to computers • Attitude to eLearning • Teacher support in eLearning • Socio-economic context • Parental support	• Access to technology and internet at home • Access to technology at school (computer labs, smart classrooms) • ICT proficiency • Access to digital content • Experience of digital content • Attitude to eLearning • Quality of district support • Use of ICT • Access to, and use of eAdmin systems • Barriers to implementing eLearning	• Commitment to eLearning • Quality of district support/ ICT integration support • Access/ use of eAdmin systems • Access to Digital content

The teacher and learner data were collected at different intervals and for different learner groups over time. Resource constraints limited the ability to collect learner data in the 63 schools for all grades. It was therefore decided to focus on a select set of grades in the sample schools:

- Grades 4 and 6 in primary schools
- Grade 8 and 10 in high schools

With the Game Changers coming to an end in 2019, data could only be collected for two years. This meant that the same set of learners could be surveyed twice in the case of the 2017 learner group: grade 4s and grade 6s would progress to grade 5 and grade 7 respectively in 2018, and similarly the 2017 grade 8 and 10s would be in grade 9 and 11 respectively in 2018. The model schools were an exception – as the 2017 model school learners were surveyed three times: twice in 2017 and once in 2018.

The 2018 learner group, also starting with grade 4 and 6 in primary school and grade 8 and 10 in high school would only be surveyed once. A tabular view of the sequencing of data collection for learners, teachers and principals are shown in table 14. The two learner groups (LG) are further explained at the bottom of the table.

Table 14: Timing of surveys: learners, teachers and principals in 2017 and 2018

Surveys: learners, teachers & principals	2017					2018					
	March	**July**	**Aug**	**Sept**	**Oct**	**Feb**	**March**	**April**	**June**	**Sept**	**Oct**
Model: learner questionnaire			LG1		LG1	LG2		LG1	LG2	LG1	LG2
Enhanced & universal: learner questionnaire					LG1	LG2				LG1	LG2
Teacher questionnaire					x			x			x
Principal questionnaire	x						x				

Notes:

1. LG1 = Learner Group 1
   2017: In primary schools, grade 4 and 6, In 2018 this group moves to grade 5 and 7
   2017: In high schools, grade 8 and 10, in 2018 this group moves to grade 9 and 11
2. LG 2 = Learner group 2
   2018: In primary schools, grade 4 and 6
   2018: In high schools, grade 8 and 10
3. Teachers completed the questionnaires three times over the two-year period, while principals completed the questionnaire twice.

Timetable schedule

Existing school timetables did not indicate whether a particular class is taught in a computer laboratory or a smart classroom (technology enabled classrooms). To establish learners' access to technology, a timetable template was developed that only the sampled enhanced and universal schools would complete (total of 47 schools). The 16 model schools were excluded as these schools had a 100% coverage due to learners having their own devices – it was therefore not necessary to establish their access to computer labs and smart classrooms.

Observations

Observations: Classroom observation schedule

A classroom observation tool developed by Microsoft was deemed suitable for the purpose of the eLearning Game Changer. A selection of schools was included for the administration of the classroom observation schedule. The maths subject advisor in every district was tasked to select two enhanced schools per district – one primary and one high school. This resulted in the set of 16 schools (two per district) not fully aligning with the 63 eLearning schools. The classroom observations were done on a quarterly basis.

Platform and implementation data

Learner and teacher level WAN usage report

The installation of probes in the full set of sample schools (111) schools enabled for WAN usage statistics to be available at a learner and teacher level.

WAN downtime report

The probes installed in the 111 sample schools provided details of when WAN was inactive (referred to as downtime).

eCulture checklist

All schools signed a letter of acknowledgement (universal schools), memorandum of understanding (model schools) or memorandum of agreement (enhanced schools). The letter sets out the technology being rolled out at the school as well as the school's role, and responsibility in the technology roll out.

CEMIS report

The CEMIS report indicated which schools had been accessing functionalities additional to the CEMIS system during the eLearning Game Changer. The additional functionalities relate primarily to the eAdmin work stream.

Technology roll-out

Technology roll-out data was already available through the Broadband School Management Information System, and other eAdmin systems.

CeI Service desk data

The CeI Service desk data was already available in the required format with some adjustments made to the CeI service desk system during the Game Changer period. One of these adjustments entailed categorising all queries into 9 categories.

ePortal usage data

The ePortal is an online educational resource that was developed by the WCED. It allows learners, teachers and other educational stakeholders to access a range of resources including video clips, audio clips, digital documents, lesson plants from anywhere. The ePortal produces generic user statistics. During the Game Changer period, work commenced to expand the login capability in order for teacher and learner level usage to be determined.

Learner and teacher data

Learner level attendance and school performance data

WCED only releases aggregated learner data, which is understandable given the concerns around the Protection of Personal Information Act and the need to protect the learner's identity. All three Youth Game Changers however aimed to impact the learner in terms of learner performance. Also, school attendance is an important predictor of learner success as chronic absenteeism is bound to impact the learner's performance at school. There was a strong drive during the Game Changer period to collect learner and teacher level data, as it enabled more detailed analysis on whether the planned interventions rendered the desired effect.

Training data for teachers

Training data for teachers and school management was available in aggregated format (school and district level). Work commenced during the Game Changer period to expand this to be recorded at teacher level, but this system would only be available after the Game Changer.

5.2.4 Data analysis

In all instances structured analyses were performed of the data using statistical methods. Excel was used predominantly as a means of analysing the data, given that most of the existing data was received in Excel format. All questionnaire data for teachers and learners were analysed using the Statistical Package for the Social Sciences (SPSS).

All teacher and learner data collection included reference to unique identifiers - CEMIS in the case of learners and PERSAL number in the case of teachers. This enabled us to compare the exact same group of people (learners and teachers) and check for the change in outcomes over the two year period. A regression analysis was also performed by a statistical expert at the DSU, whereby different learner

regression models were developed. Unfortunately, these analyses could only be conducted for learner performance data in 2018 (once implementation became more standardized). The results of the regression modelling were reported at one stocktake meeting in February 2019 only. With the closing of the DSU in June 2019, no further modelling was possible. This was unfortunate as the first results from the regression analyses suggested some positive impact on learner performance even though the Game Changer had only been implemented for three years. The summary slides on these results, as shared at the February 2019 stocktaking meeting, are included as Annexure A.

5.3 Summary

A case study design was deemed most appropriate for this study. This type of design is commonly applied in education, given its ability to capture the complexity of educational programmes. Thomas' (2011) typology of case studies was utilised to unpack the elements on the eLearning case, clarifying the following elements: the subject, object, purpose, methodological choices and approach to the eLearning Game Changer. The methodological choices were discussed in detail, explaining the selection and sampling of schools, the measurement instruments, the data collection methods and tools as well as how data analysis was performed in the eLearning Game Changer.

The WCG established the DSU in 2015 to give effect to the Game Changer policy priorities identified by the provincial cabinet. Initially, eLearning did not form part of the provincial priorities but was placed under the purview of the DSU towards the end of 2015 in response to the Premier's need for greater oversight of this programme.

I now return to the subsidiary research questions for part two to discuss in detail how the Deliverology framework was modified for the eLearning Game Changer, utilising the Deliverology steps as the analytical framework. The table below provides the structure for the rest of Part two:

Table 15: Chapter breakdown to cover the relevant Deliverology steps (and sub steps)

Deliverology step	Sub step	Chapter breakdown
Develop a foundation for delivery	Define your aspiration	Chapter 6
Plan for delivery	Determine your reform strategies	Chapter 7
	Development of indicators	Chapter 8

Deliverology step	Sub step	Chapter breakdown
	Set targets & establish trajectories	Chapter 7
Drive delivery	Establish routines to drive and monitor performance	Chapter 9

Chapter 6: Develop a foundation for delivery

In Chapter 6 and the next three chapters I follow a similar structure. In all of these chapters I commence with a description of the step as found in the "standard" Deliverology framework including a focus on the perceived shortcomings and limitations. In the remainder of each of the chapters the focus shifts to more analytical reflections and specifically a discussion of how each step was modified, as well as how – in my assessment – these modifications resulted in clear gains in the ongoing process of measuring the performance of the e-Learning Game Changer.

6.1 Define your aspiration

The first step in Deliverology is called **"develop a foundation for delivery".** As indicated in the previous chapter our specific focus is on the sub step - "Defining our aspiration". Barber et al. (2016) describe this step as follow:

> Your aspiration helps you answer the first question of delivery: "What are you trying to do?" It defines an outcome connected to your moral purpose, it defines how you'd like to see that outcome move, and it can be broken down into no more than a handful of goals – each with a metric, target and goal leader, who hold primary responsibility for achieving it
>
> (Barber, et al., 2016, p.11)

I would argue that the definition of 'defining our aspiration' as presented above is inadequate. Without more clarity on what this means, it is unlikely that any project team will reach consensus on the best way moving forward. An alternative phrasing is often used in Deliverology texts to 'clarify' what the overarching aspiration statement means, viz. "What does success look like?". It is assumed that this more easily understandable phrase (on the face of it) would assist teams arrive at a clear programme purpose. In addition, a number of steps are furthermore proposed to assist in developing a 'clear aspiration':

- Identify existing aspirations and goals taking the following into consideration: existing strategic documents, legislative mandate and political commitments/ electoral commitments
- Prioritise and refine the aspiration by selecting the most important aspirations – having 20 priorities is not feasible

- Communicate the aspiration: the goal and aspiration need to be communicated to stakeholders at all levels. The aspiration needs to be clearly formulated and well communicated as this will garner support for the work to be done.

I do agree with the Deliverology approach that the formulation of an aspiration statement does constitute one of the first steps in programme design, and that there needs to be a link between the aspiration statement and underlying "goal statements". However, I was also of the view at the time that there were at least two ways in which the Deliverology framework could be strengthened. These are: (a) clarifying the performance measurement terminology, which included a clear distinction between the output and outcomes of the intervention and (b) articulating and documenting an explicit theory of change statement for the Game Changers that introduces a distinction between short-, medium- and long-term outcomes. On the face of it, these are self-evident features of a well-designed programmatic intervention – especially if one was trained in programme evaluation theory and design. The fact of the matter is that these issues are not explicitly addressed and reflected in the standard Deliverology approach.

6.2. Modifications to step 1 of the Deliverology framework

6.2.1 Clarification of terminology

Deliverology suggests not "getting tied up by the terminology" but do agree that a "common language" must be established (Barber et al., 2016, p. 19). Despite the commitment to a common language, Deliverology's interpretation of some key concepts is lacking. The DSU decided to utilise recognised programme evaluation and performance measurement terminology in the Game Changer programmes.

A deviation from the Deliverology approach was the inclusion of outputs, with Deliverology only prescribing that milestones in support of the outcomes (or goals as they referred to it) be developed. The DSU introduced outputs to ensure clarity on what needed to be delivered on an annual basis. Milestones were developed continuously (i.e. a full set of milestones were not developed at the start of the project), and milestone timeframes were adjusted frequently making it challenging to determine on the basis of the milestones only whether the desired outcomes were going to be achieved. The outputs bridged that gap. The South African public sector has been utilising performance measurement and programme evaluation terminology for more than a decade, and officials were therefore familiar with the language of outputs, outcomes, performance indicators etc. The distinction between outputs and outcomes was not necessarily well understood by the officials but utilising the same terminology did assist.

Separating outputs from outcomes become important when the performance monitoring is undertaken as will be demonstrated in Chapter 9.

The textbox below provides a definition of the terminology used in the Game Changer programmes:

Performance measurement/ programme evaluation/Deliverology terminology

Goal: A statement, usually general and abstract, of a desired state towards which a programme is directed

Objective statement: Specific statements detailing the desired accomplishments of a programme

Outcome: is the state of the target population or the social condition that a program is expected to change (typically knowledge, attitudes and behaviour)

Output: the goods and services produced by an intervention

Milestones: Milestones are signposts through the course of your project, ensuring you stay on track

(Rossi et al, 2004; Mouton, 2019, Teamgantt, 2020.)

Figure 18: Performance measurement terminology in the DSU

The links between these concepts are depicted graphically in Figure 19. Level 1 (denoted by the "1") marks the starting point with the identification of a goal statement, followed by the formulation of more specific objective statements for the various work streams (denoted by "2"). The outcomes for each of the work streams reflect the change that needs to take place in the target group (denoted by the "3"). To achieve these outcomes, products or services get produced during a programme - referred to as outputs (denoted by the "4"). The final element of the plans entailed the development of milestones, which provided signposts as to whether the outputs were on track or off track (denoted by the "5").

The alignment with programme evaluation and performance measurement terminology is indicated to the right of the figure noting the deviations. Even though the terminology of goals and objectives is found in government documents – commonly referred to as strategic goals or strategic objectives, there is a tendency to conflate the terminology. Within this study our understanding of goal and objectives aligns with standard programme evaluation practice.

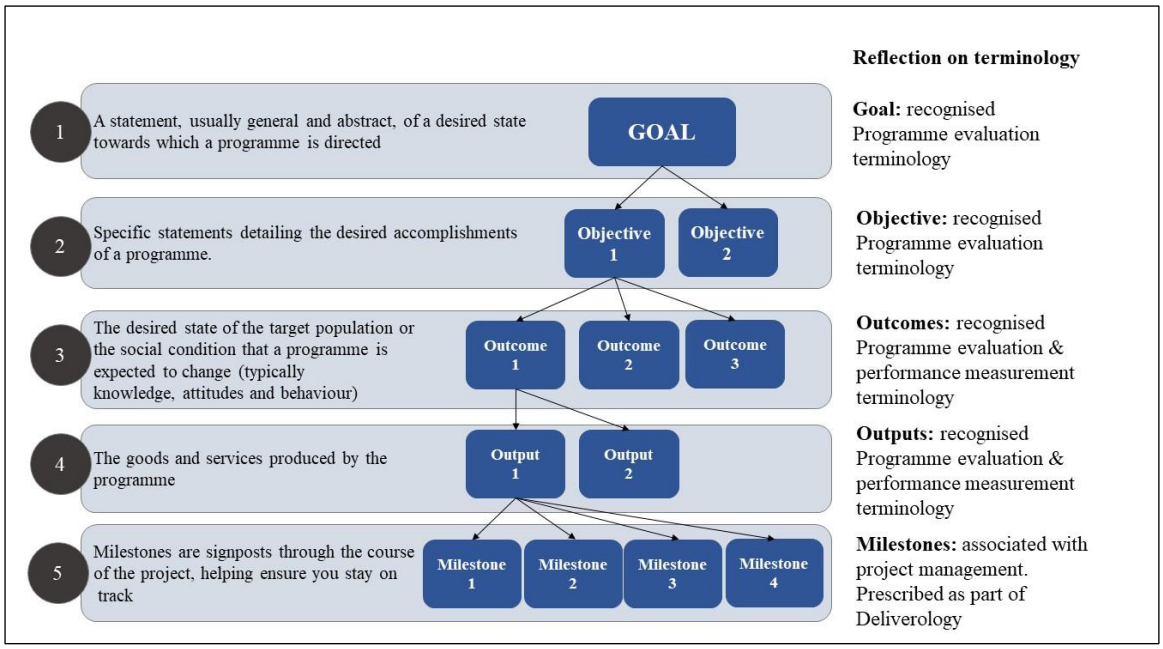


Figure 19: The logic of the Game Changers and terminology utilised

(Mouton, 2018, utilising terminology from Mouton, 2019, Rossi et al, 2004)

Starting with the top level, the eLearning Game Changer goal statement was informed by the e-Education vision (dated 2012), which stipulated a two-fold purpose for e-Education: improved learner outcomes and better prepared learners for the 21st Century by 2030. The 'problem' that the eLearning Game Changer was addressing is the long-standing sub-standard learner performance in addition to many learners being ill-equipped for the ICT demands of the 21st century.

Given that the Game Changer programme only had three years of implementation it was decided that the eLearning Game Changer needed to focus on what could be achieved over this relatively short period of time. After much deliberation with WCED, it was agreed that the eLearning Game Change would focus on enhancing teaching and learning. A representation of this goal statement in relation to the e-Education vision purpose statements is depicted in Figure 20. Ideally one would hope to see an improvement in learner outcomes because of eLearning. It was deemed to be unrealistic to have a goal around improved learner outcomes given the short time span of the programme.

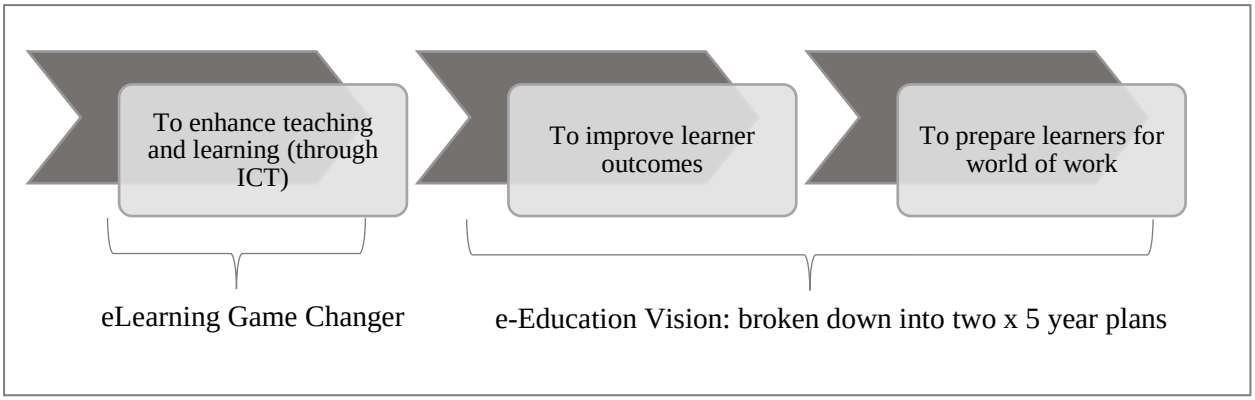


Figure 20: eLearning Game Changer within the context of the e-Education vision

Mathematics and language performance remain a grave concern for the WCED, and the country. Despite slight improvements in systemic test results for grades 3, 6 and 9 every year, absolute performance remained at low levels. Another concern was the continuing downward trend in mathematics and language results as learners' progress from grade 3 to grade 6 to grade 9. The WCED had prioritised the improvement of mathematics and language performance in all their strategic documents and annual targets as reflected in their statutory planning document, the annual performance plan. The WCED has also developed a mathematics and language strategy which sets out the various programmes being undertaken to address the sub-standard performance. It made sense to align the eLearning Game Changer to these existing initiatives and the eLearning goal statement was therefore further refined to reflect this focus as follows:

To enhance the teaching and learning experience predominantly in Mathematics and Language through the use of ICT

The focus on mathematics and language added more complexity to the implementation of the eLearning Game Changer. For example, many subjects other than mathematics and language lend themselves better to the development of stimulating and interesting digital content (for example, Geography and History). Similarly, teacher training and development was already taking place across the board, with these teachers covering multiple subjects. It was therefore agreed to only focus on mathematics and language where it made sense to do so – for example, when rolling out technology enabled classrooms (referred to as smart classrooms), schools were requested to prioritise access to these classrooms for mathematics and language subject teachers.

Setting the boundaries for mathematics and language also needed to be done, given that the Western Cape province has three official languages, namely English, Afrikaans and isiXhosa. It was decided that mathematics would encompass mathematics[11], pure mathematics and mathematics literacy. In the case of language, the focus would be on English first additional language and English home language as this would cover all learners.

6.2.3 Developing a theory of change that distinguishes between short, medium and long term outcomes

The core principle of theory-based approaches to programme evaluation, i.e. the construction of a theory of change, was utilised during the programme design phase of the Game Changer programmes. Some authors (English, Cummings & Straton, 2002; Owen & Rogers, 1999) refer to this phase in programme design as 'clarificative evaluation'.

> Clarificative evaluation clarifies the underlying rationale of a program. Program developers use this information to think through and make explicit the logic that supports the program, including assumptions about how its components link to produce the desired outcomes. (English et al., p. 126).

Theory based evaluations emerged around the 1980s to counter the "black box" approach to programme evaluation. As Chen (2005, p.232) states: "It (theory-based evaluations) can provide stakeholders with an understanding of whether a program is reaching its goals and document insightfully the how's and why's of program success or failure". This statement by Chen already suggests that theory-based evaluations may include two kinds of 'programme theories': theories that answer 'why' questions and theories that answer 'how' questions. In the diagram below, Wildschut (2019) follows this convention and sides with the Funnel and Rogers (2011) point of view.

[11] All learners take Mathematics until Grade 8. There after Mathematics becomes a subject choice where learners can choose to either continue with Pure Mathematics or Mathematics Literacy

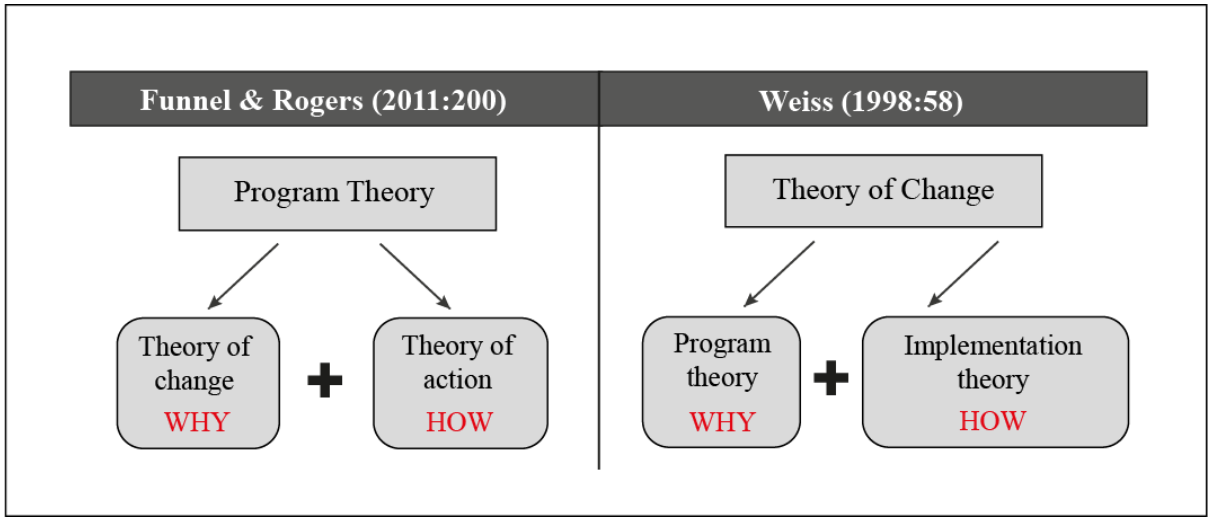


Figure 21: Different interpretations of programme theory *vis a vis* theory of change

A theory of change focuses on the causal pathways of outcomes, capturing an underlying hypothesis of how change will come about. As Dhillon and Vaca (2018, p.65) explain "the theory of change, then, becomes a roadmap, providing pathways of outcomes that lead to the organisational mission." A theory of change has many advantages: firstly, it provides strategic clarity and ownership around the purpose of a project or programme, and secondly, it allows for the hypothesis to be tested and then be adjusted should the causal pathway prove faulty (Dhillon & Vaca, 2018). In the eLearning Game Changer, I followed this use of the term 'theory of change'.

However, as far as the second 'sub-component' of the programme theory is concerned, I preferred to use the 'logic model framework' rather than a 'theory of action'. This decision was informed more by a pragmatic consideration than any strong theoretical or conceptual consideration.

I found that it was easier to explain the logic and steps in the implementation of the Game Changers to the different stakeholders who were involved in the delivery of the intervention when using the tabular format of the logic model. I agree with Wildschut and others that the logic model approach presents a static view of how implementation is supposed to unfold. However, I believe that this 'limitation' is outweighed by the fact that the tabular format forces one to be very clear about the 'horizontal' alignment between inputs (resources), activities and outputs (delivery) and outcomes (achievements) at each intervention level and the 'vertical alignment' between the intervention levels (which captures any conditionalities embedded in the implementation process). Given the complexity of the eLearning Game Changer, the 2016 logic model consisted of 78 rows. To translate each row in the logic model

into a narrative theory of action would not be feasible. I would maintain that any attempt to present this tabular format as a 'theory of action' or 'implementation theory' would have been less successful.

As indicated in Chapter 5, work on the eLearning programme was well underway by the time the Game Changer programme started. The e-Education vision sets out six objective statements (WCED, 2015, p.3):

1. Teaching: Empowering teachers and education managers to use technology effectively and efficiently.
2. eLearning: Extending learning opportunities and access to learning resources anywhere and anytime.
3. Curriculum: Provisioning of open access learning material, objects in various formats i.e. video, animations, simulations, collaborations, quizzes etc.
4. Systems: Robust and reliable ICT systems that support e-Education.
5. Environment: Access to connectivity is enabled throughout the school, i.e. inter-connectivity to other WCED sites, as well as through internet connectivity.
6. e-Administration: Reducing manual administration and providing business intelligence for planning and management.

When the eLearning programme was 'upgraded' to a Game Changer one of my first tasks, as the performance tracking director, was to initiate a close working relationship with the eLearning team in WCED to undertake the design of the Game Changer programme. For this I applied clarificative evaluation principles.

Bearing in mind the goal statement, the first step was to refine the objective statements. Through this process a new configuration of the e-Education strategy emerged: six work streams were identified, and the objective statements were subsequently aligned to these six streams. Table 16 below presents the revised objective statements and their alignment with the six work streams:

Table 16: The objectives per work stream

Work stream	Objective statements
eInfrastructure	To provide all schools in the Western Cape with access to reliable, well managed connectivity, related infrastructure and support systems
eTechnology	To provide all schools in the Western Cape with access to teacher and learner technology

Work stream	Objective statements
eTeachers/ eOfficials	To equip principals, teachers, school management support and curriculum teams to use technology effectively and innovatively To support principals, teachers, school management support and curriculum support teams to use technology effectively and innovatively
eContent	To provide up to date digital content that is responsive to the needs of teachers and learners
eAdmin	To reduce teachers' and principals' manual administration towards more effective planning and management
eCulture	To create an enabling environment for the roll-out of eLearning through strong leadership and a shared eLearning culture within WCED and Western Cape schools

(DSU, 2016a)

With the objectives finalised, outcome statements were formulated for each work stream. This was done by way of extensive workshops, ensuring representation from all relevant stakeholders. This process of developing proper outcome statements was time consuming but allowed the WCED and CeI to think through the change they would like to bring about, during the Game Changer period – and also beyond.

The table below expands on the previous table by including the outcomes per work stream. The outcome statements have been condensed for the purpose of brevity in that it covers multiple target groups (schools, teachers, and learners) and constructs (for example support, aware, use, integrate etc.). In the logic model, the outcome statements are broken down to only cover one target group and one construct. The table below also contains the three learner level impact statements, which reflect the longer term results of the eLearning programme provided a) the programme is implemented as planned and b) all other outcomes are achieved.

Table 17: eLearning outcomes per work stream

Work stream	Objective	Outcomes and Impact statements
eInfrastructure	To provide all schools in the Western Cape with access to reliable, well managed connectivity-related infrastructure and support systems	Schools and teachers have better connectivity
		Schools receive better technological support
eTechnology	To provide all schools in the Western Cape with access to teacher and learner technology	Learners and teachers have better access to technology

Work stream	Objective	Outcomes and Impact statements
eOfficials	To support principals, teachers, school management and curriculum teams to use technology effectively and innovatively	Teachers and principals are aware of the ICT integration support systems in place at districts
		Teachers and principals receive ICT integration support from district
		Curriculum support officials provide better ICT support to teachers and principals
	To equip principals, teachers, school management team and curriculum teams to use technology effectively and innovatively	Teachers use available technology in their teaching practices
		Teachers integrate available technology into their teaching environment
e Content	To provide up to date digital content that is responsive to the needs of teachers and learners	Learners and teachers have better access to digital resources
		Teachers integrate available digital resources into their teaching environment
eAdmin	To reduce principals' manual administration towards more effective planning and management	Targeted principals have access to, and use the digital school admin systems and school dashboards
eCulture	To create an enabling environment for the roll out for eLearning through strong leadership and a shared eLearning culture within WCED and Western Cape schools	School principals support the eLearning GC
Learner Impact Level		Learners have enhanced motivation to learn
		Improved learner outcomes
		Increased classroom participation by learners

In May 2016, an external review was conducted of the eLearning programme design. Several tasks had to be performed as part of this review, one being the development of a theory of change. Utilising the work done so far on formulating the outcomes, the following theory of change was produced for the eLearning Game Changer:

IF

Principals support the eLearning Game Changer

AND

Schools and teachers have better connectivity

AND

Learners and teachers have better access to technology (smart classrooms, learning devices)

AND

Schools receive better technological support

AND

Teachers and principals are aware of, and receive the ICT support available from the district

AND

Curriculum support officials (including planners) provide better ICT support to teachers and principals

AND

Learners and teachers have better access to digital resources

AND

Principals have access to digital school admin systems and school dashboard

AND

Teachers use available technology in their teaching practice, and integrate available technology and digital resources into their teaching environment

AND

Principals use the digital school admin systems and school dashboards

THEN

Learners will be more motivated to attend school, learners will be engaged during classroom teaching and Learner performance will improve

The colour coding represents the distinction between short-, medium- and long-term outcomes. The yellow sections constitute the short-term outcomes, while the blue denotes the medium term outcomes. The green represents the long-term outcomes and impact of the eLearning programme.

Formulating the theory of change had two benefits: firstly, it provided the WCED and CeI project staff with the opportunity to explicate the eLearning programme theory. Before the Game Changer, only objectives and a broad purpose statement were formulated not the causal pathway of how change would happen. Through the development of a theory of change a clear distinction was made between the short, medium and longer term outcomes. This allowed the work of the Game Changer to become very

focused, recognising that the shorter term outcomes around access needed to receive the most emphasis as it constituted the building blocks for the medium and longer term outcomes.

Secondly, the theory of change guided the indicator selection and formulation phase, as well as the ensuing data collection for all the outcome statements. In this way the theory of change would be validated or rejected: the data would show whether the achievement of all these outcomes ultimately lead to the three learner impact statements (learners being more motivated to attend school, learners being more engaged during classroom teaching and ultimately improving learner performance). If not, it meant the theory of change was faulty and needed to be modified. This is the premise of theory-based evaluations – data collection is not done in a haphazard manner but with the purpose of validating a causal theory in order to enhance the body of knowledge in a particular discipline.

6.3 Summary of modifications to step 1 and the gains produced

Proponents of Deliverology agree that terminology matters but do not always define the performance measurement concepts clearly. In addition, the concept of 'outputs' is not utilised in Deliverology; only milestones which are developed in support of goals. The first modification made to the Deliverology framework led to the inclusion of outputs – a modification which proved to be essential later when monitoring the performance of the intervention.

Similar to the field of programme evaluation, Deliverology prescribes as a first step the identification of the programme aspiration. In addition, there is an implicit undertaking to programme logic during Step 1 of Deliverology given that the "goals" must be selected in support of the overall aspiration. However, no distinction is made between short-, medium-, and longer- term outcomes in Deliverology. The second modification entailed performing a clarificative programme evaluation during the design phase of the eLearning Game Changer to strengthen step 1 of Deliverology. This constitutes two aspects: the development of a theory of change and a logic model. In developing a theory of change, the eLearning Game Changer programme design was strengthened in two ways: firstly, the causal pathways were made explicit and secondly it provided a clear distinction between short-, medium-, and longer- term outcomes. In terms of measurement, the theory of change would eventually guide the selection and formulation of indicators for the eLearning Game Changer.

With this in mind, I revise the wording for step 1a as follow:

From:

Define your aspiration

To:

Define your **goal, outcomes (short, medium and long term using a theory of change) and outputs**

Chapter 7: Plan for delivery

7.1 Develop reform strategies

With the overall aspiration (goal statement) and outcomes finalised, work on a detailed delivery plan commenced. In the Deliverology framework, Step 3 deals with the development of the delivery plan:

> Your aspiration answered the first question of delivery: What are you trying to do? This part of the book answers the second: How are you planning to do it? Specifically, what's your delivery plan for achieving the goals you've set (Barber et al. 2016, p.110)?

Deliverology does not prescribe fixed content for the delivery plan but proposes that the following form part of the delivery plan: i) reform strategies, ii) delivery chains and iii) targets and trajectories (Figure 22). These also constitute the three sub-steps of step 3, of which only the reform strategies are dealt with in this chapter. The targets and trajectories are discussed in the subsequent chapter.

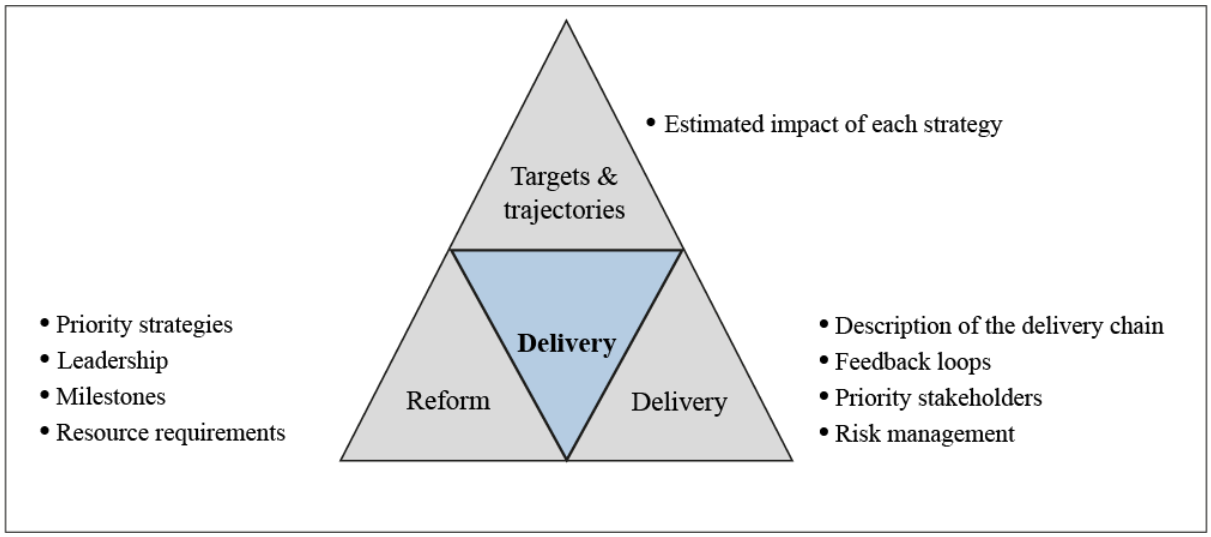


Figure 22: Proposed content of a Delivery plan

(Barber et al., 2016, p.119)

The reform strategies (left part of the triangle) are defined as a "deliberate and coordinated set of activities that is designed to help you achieve one or more of your outcome goals. Together, the strategies you prioritise and implement constitute your overall reform strategy" (Barber et al. 2016, p. 127).

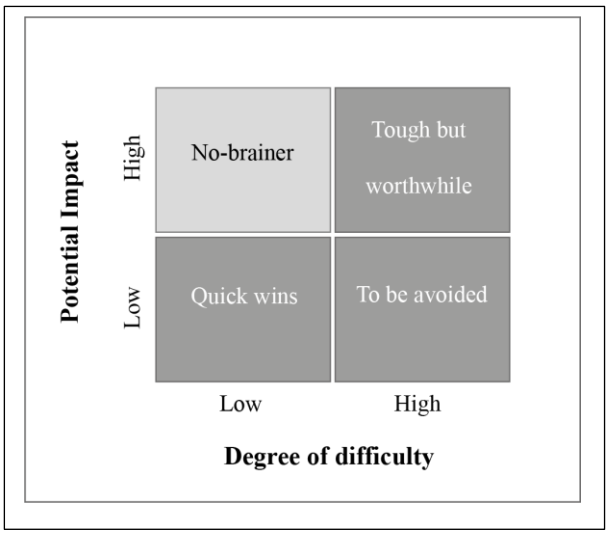


Figure 23: Prioritisation matrix

In terms of prioritising potential strategies, Deliverology proposed a very useful matrix which assists in deciding which strategies to pursue. The adjacent figure (Barber et al., 2016, p.134) contains the prioritisation matrix, as well as some cues on the placement of projects as per the four blocks.

The elaboration of these reform strategies entailed the development of critical path milestones to ensure delivery can be tracked in the necessary detail, viz. ensuring every strategy has a clear leader (not just the name of an organisation) as well as the necessary resources in place. A famous quote used often to drive home the importance of budget is a comment made by Idris Jala, the former head of the Malaysian Delivery Unit (PEMANDU): "A strategy without a budget is just a plan".

7.2. Modifications to step 3 of the Deliverology framework

7.2.1 Development of reform strategies using the logic model

The delivery plan represents the overarching plan for the eLearning Game Changer. The DSU opted to call their delivery plans "roadmaps". My motivation for utilising the logic model in developing our reform strategies, as contained in the roadmap, was based on the coherence this tool offers. The logic model addresses not only the links between outcomes and activities but also specifies the indicators and targets (where appropriate). It therefore pulls together the reform strategy elements as well as measurement issues (specifically targets) as set out in Figure 22. As alluded to above, Deliverology recommends the use of a "theory of action", which presumably is best understood as an 'implementation theory'. However, the explicit introduction of logic model categories represents an expanded version of a theory of change which shows how the various pieces are aligned to each other.

The roadmap was a formal document, signed off by the provincial cabinet. It contained the rationale for the eLearning Game Changer, the goal statement, some elements of the logic model, indicators, governance structures and a communication plan. More detail on each of the sections of the roadmap is provided below:

- Problem statement: what specific problem is the Game Changer attempting to address
- Goal statement: what is the overall intent of the Game Changer over the three years
- Work streams: in what areas must we intervene to bring about the desired change as set out by the three-year goal statement?
- Outcomes per work stream: what needs to change, for whom, in each of the work streams?
- Annual outputs: what products or services must be produced to give effect to the various outcomes?
- An annexure dedicated to the performance tracking of the Game Changer, covering the theory of change, indicators to measure the Game Changer as well as targets where available.
- An annexure setting out the communication strategy – mainly externally focused, i.e. communication to support delivery.
- An annexure explaining the governance structure of the Game Changer.
- An annexure that contains the delivery chains that were developed.

The eLearning logic model was divided into two sections, which helped officials to distinguish the two major parts of the logic model: “What we need to do” and “What we need to achieve”. The table below shows the elements of the two sections, with the blue highlights denoting the typical elements found in a logic model. The left side of the table (What we need to do) included the outputs, targets, output indicators, milestones, person responsible and timelines while the right side of the table covers the outcome related columns of a logic model (outcomes, targets, outcome indicators, source of evidence). The “non typical logic model” elements, prescribed by the Deliverology approach are milestones, assigning a person responsible and attaching timelines. This is in response to the suggestion that reform strategies must be “built out” to include critical path milestones and have clear timelines and roles and responsibilities attached – these are the elements that focus on the actual delivery. Given the emphasis on data in Deliverology, it was decided to consider the availability of data early in the process, and hence a column was added showing whether there was existing data in support of the indicator.

Table 18: (Expanded) elements of the logic model as developed in the DSU

WHAT WE NEED TO DO	WHAT WE WANT TO ACHIEVE
○ Target group ○ Milestones ○ Timelines ○ Person responsible (for milestone) ○ Outputs, with targets broken down per annum ○ Output/ Performance Indicators ○ Existing data (Yes/No)	○ Outcome statements ○ Target (where applicable) ○ Outcome indicators ○ Source of evidence ○ Existing data (Yes/No)

The table below contains an extract from the 2016 roadmap document, showing how the logic model elements of the eInfrastructure and eTechnology work streams, i.e. the objectives, outcomes, outputs and targets were ultimately presented. The objective statements for the two work streams are shown at the top of the table, followed by the outcome statements for the two work streams and ultimately the outputs for the three categories of schools over the three year Game Changer period[12]. Without the logic model it would have been challenging to derive at a coherent view of the work streams:

Table 19: eInfrastructure and eTechnology work streams

eInfrastructure and eTechnology						
Objectives: To provide all schools in the Western Cape with access to: - reliable, well managed connectivity, related infrastructure and support systems; - teacher and learner technology						
Outcomes: - Schools and teachers have better connectivity - Schools receive better technological support - Learners and teachers have access to technology (learner devices and smart classrooms)						
Category	**WAN**	**LAN**	**Laboratory refresh**	**Wireless access point**	**SMART classrooms (Teacher devices)**	**Learner devices**
Universal schools	y	x	Y	y	x	x
Enhanced schools	y	y	Y	y	y (40-50%)	x

[12] Targets shown for financial years. Even though schools follow a calendar year, budgets are released according to financial years (Financial year stretches from 1 April to 31 March)

eInfrastructure and eTechnology						
Model schools	y	y	Y	y	y (100%)	y
2016/17	**What Does Success Look Like (Annual Targets)**					
Universal schools	All	-	388	388	-	-
Enhanced schools	All	162	75	75	1 200	-
Model schools	All	8	16	16	200	8000
2017/18	**What Does Success Look Like (Annual Targets)**					
Universal schools		-	388	388	-	-
Enhanced schools		100	75	75	1 000	-
Model schools		8			200	8000
2018/19	**What Does Success Look Like (Annual Targets)**					
Universal schools		-	388	388	-	-
Enhanced schools		100	75	75	1 000	-
Model schools						

(DSU, 2016a)

During the logic model development process, indicators were also identified for all the outputs and outcomes. This ensures the indicators are fully aligned to what needs to be measured. The indicators were included as an annexure to the roadmap. The table below lists all the output and outcome indicators per work stream. The output indicators have been shaded in grey.

Table 20: List of output and outcome indicators per work stream

Work stream	Output & Outcome statements	Performance indicators and outcome indicators
eInfrastructure	1278 schools have WAN	Number of schools connected to Broadband (per school category)
	366 enhanced schools have LAN	Number of enhanced schools with LANs
	972 Schools with wireless access points	Number of schools with wireless access points as part of the slim lab rollout (universal and enhanced schools)
	School ICT queries tracked (WCG and non-WCG related)	Number of queries logged at CeI service desk (per category of query)
	Teachers have better access to internet connectivity	Number of teachers that connect to wireless access points to enable own device (per school category)
	Schools have better access to internet connectivity	Data being accessed and downloaded by schools (by type, per school category)
		Schools' data volumes (per school category)
	Schools receive better technological support	Number of queries reported to CeI service desk that are resolved within five working days (per category of query, per school category)
eTechnology	1160 schools have slim laboratory refreshed	Number of schools with slim laboratory refresh (universal and enhanced schools)
	7530 smart classrooms deployed with teacher devices	Number of smart classrooms (teacher devices) deployed (model and enhanced schools)
	16183 devices distributed to model schools	Number of learner devices that has been distributed to model schools
	Learners have improved access to technology	Percentage of maths & language contact time spent in smart classroom environment (For mathematics & language, per grade, per school category)
		Percentage of maths & language contact time spent in laboratory environment (For mathematics & language, per grade, per school category)
		Number of model school learners that connect devices on a daily basis

Work stream	Output & Outcome statements	Performance indicators and outcome indicators
eTeachers/ eOfficials	32214 Principals and teachers trained in basic ICT competencies	Number of school staff trained in basic ICT competencies (principals and teachers; per school category)
	4056 teachers trained in ICT integration	Number of teachers trained in ICT integration (per school category)
	1239 Principals and school management officials trained in ICT integration	Number of school management officials trained (Principals, deputy principals, HoDs; per school category)
	Teachers are aware of the ICT integration support system in place at districts	Number of teachers that are aware of the ICT Integration support available to them (per school category)
	Principals are aware of the ICT integration support system in place at districts	Number of principals that are aware of the ICT Integration support available to them (per school category)
	Principals receive ICT integration support from district	Number of principals that request support from district staff (per school category)
	Teachers receive ICT integration support from district	Number of teachers that request support from District staff (per school category)
	Curriculum support officials provide better ICT support to principals	Number of principals that are satisfied with support they receive from the district (per school category)
	Curriculum support officials provide better ICT support to teachers	Number of teachers that are satisfied with support they receive from the district (per school category)
	Teachers use available technology in their teaching practice	Number of teachers that use wireless access points to enrich educational practices (per school category)
		Number of teachers at model schools that use learning management systems to plan and deliver online lessons
	Teachers integrate available technology into their teaching environment	Number of teachers that integrate ICT into their teaching environment (per school category)

Work stream	Output & Outcome statements	Performance indicators and outcome indicators
eContent	Five digital resources available for every Maths and Language topic grade 4-12	Number of topics with digital resources available (per subject, per grade)
	Learners have better access to digital resources	Number of learners that indicate they choose digital resources as a first choice when learning (per school category)
		Count of CAPS aligned downloaded/ page views by learners (per subject, per grade, per school category)
	Teachers integrate digital resources into their teaching environment	Number of teachers that access professional learning communities on e-Portal and through other collaborative platforms on a recurring basis (per school category)
		Number of teachers that upload content onto e-Portal (per school category)
		Number of teachers that share resources with other teachers through social media and electronic communication (per school category)
eAdmin	16 model schools, 50 enhanced schools, 25 universal schools have web-based school admin systems	Number of schools with web-based school administration system in place (per school category)
	16 model schools, 150 enhanced schools, have data dashboards	Number of schools where data dashboards have been made available/ rolled out (model and enhanced)
	16 model schools, 10 enhanced schools have learning management systems	Number of schools with learning management system implemented (model and enhanced)
	16 model schools, 10 enhanced schools have parent/ learner portal implemented	Number of schools with parent/ learner portal implemented (model and enhanced)
	Targeted principals have better access to digital school admin systems and school dashboards	Number of schools that upload learner assessment results to a central repository (model and targeted enhanced)
		Number of schools that access learner assessment and attendance data on the dashboard (model and targeted enhanced schools)
	Targeted principals use the digital school admin systems and school dashboards	Number of schools that use a school admin system to collect school based assessment and attendance data (model and targeted enhanced)

Work stream	Output & Outcome statements	Performance indicators and outcome indicators
		Number of school principals that report improved data collection and management (model and enhanced)
eCulture	1499 schools sign Memorandum of understanding (MOU)/ memorandum of agreement (MOA) or Letter of commitment	Number of schools that sign MOU/MOA/ Letter of commitment (per school category)
	School principals support the eLearning GC	Number of school principals that express commitment to eLearning Game Changer (per school category)
Learner Impact Level	Enhanced motivation to learn	Number of learners attending class (per grade, per school category)
	Improved learner outcomes	Learner performance scores (mathematics & language; per grade, per school category)
	Increased classroom participation	Number of learners who actively participate in classroom activities (per selected grades)

In the next chapter I will expand on the indicator development in more detail.

7.3 Summary of modifications to Step 3 and the gains produced

Step 3 of Deliverology gets into the detail of planning, making some useful recommendations around the development of delivery plans. A major advantage of the Deliverology approach is the variety of practical tools available to guide one through the various steps. Delivery chains, the prioritisation matrix and the use of trajectories are just some of the tools available during step 3 to navigate the development of a sound delivery plan. As part of the delivery plan reform strategies must be developed. The reform strategies must show the links between the programme activities and "outcome goals".

As argued above, I utilised the logic model framework to represent the links between the different elements of the intervention. The eLearning roadmap extracted for the eTechnology and eInfrastructure work streams demonstrated the "product" of the logic model. The gains from utilising the logic model are threefold: firstly, one can see the links between the outputs and outcomes and secondly, the targets which were set as derived from the logic model. The logic model not only proved to be useful in organising all the programme elements (i.e. outcomes, outputs), into a coherent view of the logic of the eLearning Game Changer but also guided the indicator selection process by ensuring a more careful consideration of the type of indicators best suited to measure the theory of change. This third benefit will be further discussed in the next chapter.

The revised wording for Step 3, taking into consideration the use of the logic model is indicated in bold font.

From:

> Step 3: Plan for delivery. This includes:
>
> - Sub step 3.1: Determine your reform strategy which include priority strategies, milestones, people responsible and the resource requirements.
> - *Sub step 3.2: Draw the delivery chain (not covered in this study)*
> - Sub step 3.3: Set targets and establish trajectories

To:

> Step 3: Plan for delivery.

- Sub step 3.1: Determine your reform strategy **(addition of logic model for performance monitoring)**
- *Sub step 3.2: Draw the delivery chain (not covered in this study)*
- Sub step 3.3: Set targets and establish trajectories

Chapter 8: Expanded step: conceptualise and develop the indicators

8.1 Introduction

Deliverology offers some direction as far as the selection and formulation of indicators are concerned but does not deal with this function in detail. Deliverology also encourages that targets and trajectories be developed for all (or most) goal statements (ref. top triangle, Figure 22).

In this chapter, I begin with a brief overview of indicator development and target setting as found within the "standard" Deliverology framework. This is followed by a more in-depth discussion of the necessity of having appropriate indicators for performance measurement as addressed in the eLearning Game Changer.

8.2 Indicator development as well as target setting

I provide a brief overview of the performance measurement aspects that Deliverology addresses, which includes i) guidance on metrics and ii) target setting practices. I also consider Deliverology's approach to data collection.

Deliverology primarily utilises the terminology of "metrics". A goal metric is defined as follows: "a type of data or measure that you use to measure progress against your goal" (Barber et al., 2016, p.166). The problem with this definition is that it does not properly tell us what a metric *is*, but rather what the metric *does* – in this example measuring progress towards a goal. This is referred to as definitional fallacy where the definition covers the purpose or functions of the concept instead of describing the nature of concept. I return to this issue below.

Deliverology suggests the use of different categories of metrics to ensure different aspects of performance is measured. The goal metric is commonly referred to in Deliverology as key performance indicators (KPIs). The use of KPIs is commonly found in the business community (UNAIDS, n.d.) which underscores the business influence in Deliverology. Three other metric options are described in the Deliverology literature:

- Progress metrics: also described as leading indicators, assist in predicting future performance. Leading indicators are associated with the economy and are used to predict economic trends (UNAIDS, n.d.).

- Process metrics: provides a view on "whether the work we're doing to impact the goals is happening" (Barber et al., 2016, p.73). This aligns to output indicators in programme evaluation as it aims to capture progress towards implementation.
- Perverse metrics: these metrics are geared at early identification of any unintended consequences that happen because of the project or programme strategies.
  (Barber et al., 2016, p. 72-73)

Aside from the suggestion to use different categories of metrics very limited guidance is offered in terms of how to select or develop new indicators. In Step 2 of Deliverology, three other criteria for selecting indicators are provided: "prioritise the ones that are clearly linked to your aspiration, easier to collect, and actionable" (Barber et al., 2016, p.73). Even though this advice stipulates the need for alignment between the indicators and the goals, it could lead to a bias towards the easily measurable indicators.

Deliverology strongly suggests that a baseline and target be set for all goals, and that trajectories be developed on this basis. The target constitutes the numeric end value to be achieved, while the trajectory breaks this target down into interim targets, i.e. annual or quarterly targets to ensure any deviation is identified early on. Target setting is commonly done against some form of benchmarking, i.e. "comparing data in a way that helps you get a sense of what performance is versus what it should be" (Barber et al., 2016, p. 74).

In terms of roles and responsibilities: a delivery unit would typically assist with the functions described above, i.e. selecting appropriate metrics, advising on target setting and making recommendations around innovative data collection strategies but the execution of all data collection activities reside with the lead department. This means that with any new data collection the lead department would need to develop the data collection instruments, put the system and processes in place to collect the data, capture and analyse the data. The delivery unit would step in again with the visualisation of the data – assisting with the production of impactful visuals as part of the stocktake reports. Because the delivery unit was not directly responsible for the data collection, it would require evidence of the stated performance to be provided and the delivery unit could even occasionally collect their own data at a small scale to confirm the departmental findings. But in the main, the delivery unit was not responsible for the continuous collection of data.

The aim in this chapter, then, is to articulate some of the measurement aspects in more detail since it is not fully explicated in the Deliverology approach.

8.3 Modifications to the Deliverology framework: introducing an expanded step on indicator development and the DSU approach to target setting

I proposed and included an additional sub-step as part of Step 3 for the following reasons: Deliverology falls short in defining indicators properly, and secondly, even though there is an intent to utilise a range of indicators, the selection is not necessarily guided by a framework which assists with the indicator selection process. In the previous two chapters, I demonstrated how the principles of clarificative evaluation were incorporated into the steps of Deliverology. The identification and selection of indicators, I argued, followed logically from this, and hence were included in the final, detailed logic model.

In terms of target setting, measuring outputs is relatively straightforward as these "products" or "services" are tangible and can therefore be counted and quantified with little effort. The challenge is setting targets for "customised" short and medium term outcomes as per the theory of change. The lack of a baseline prevented any meaningful target setting. The decision was therefore taken to only set targets at the output level, and to use the Game Changer period to establish a baseline for the outcomes. No outcome targets were therefore set.

In addressing the shortcomings of Deliverology around indicator development I commence with a definition of indicators before moving to a categorisation of indicators. For this I draw on the field of programme evaluation as well as the social indicator movement. The social indicator movement is relevant to this chapter as much of the terminology associated with indicators has its roots in the social indicator movement.

8.4 Indicator development and formulation

8.4.1 Defining indicators

In clarifying what an indicator is, the definition from Carley (1981, p.2) provides a good starting point:

> This points up two important characteristics of social indicators: they are surrogates and they are measures. As surrogates, social indicators do not stand by themselves. Rather than translate abstract or unmeasurable social concepts, like safe streets into operational terms that allow

> consideration and analysis of the concept, like number of crimeless days. A social indicator must always be related back to the unmeasurable concept of which it is a proxy. And as measures social indicators are concerned with information which is conceptually quantifiable and must avoid dealing with information which cannot be expressed on some ordered scale.

As shown by Carley, indicators are 'referential measures' which means that they refer or point to something beyond the directly observable or visible. Kaplan (1964, as cited in Babbie and Mouton, 2001, p.110) in his categorisation of "things scientists can measure" distinguishes between things that can be directly observed compared to indirectly observables and constructs. Kaplan defines constructs as "theoretical creations based on observations but which cannot be observed directly or indirectly. IQ is a good example. It is constructed mathematically from observations of the answers given to a large number of questions on an IQ test".

It is for this reason that I intentionally use the word indicator and not metric or measure in this study. With social interventions, the outcomes one aims to achieve tend to be intangible – they are often couched in abstract terms as concepts or constructs. They can often not be directly observed, making the measurement thereof challenging. At the very best one can get an approximation, or indication of the achievement of outcomes.

In dealing with these intangible and unobservable concepts two processes are undertaken as part of social measurement: conceptualisation and operationalisation. Conceptualisation is the process one follows to derive at a common understanding of the meanings of the key concepts (or constructs) we aim to measure. Take for example the construct of "ICT integration": without a shared understanding of what this means, one is unable to proceed to the next step in the measurement process, namely operationalisation. "Operationalisation is the development of specific research procedures (commonly called operations) that will result in empirical observations representing those concepts in the real world" (Babbie and Mouton, 2001, p.128). This links with Carley's second characteristic of social indicators namely the need to develop an operational definition of how the construct will be measured, i.e. the indicators.

The process of operationalisation ultimately results in the "unpacking" of abstract concepts to a point of understanding whether an empirical measure (a single or composite indicator) is sufficient to generate data about the concept (Babbie and Mouton, 2001). Conceptualisation is therefore about clarifying

what we will measure, whereas operationalisation is concerned with how we will measure the selected concepts/ constructs.

8.4.2 Categorisation of indicators

Indicators can be categorised in different ways, with different disciplines using different typologies and terminology. For the purpose of classifying the eLearning Game Changer indicators, I will draw on contributions from both the social indicator movement and the programme evaluation tradition. This results in three classifications of indicators:

- A classification that distinguishes between objective and subjective indicators (which corresponds with statistical levels of measurement and the distinction between quantitative and qualitative).
- A classification that takes the complexity of the construct to be measured into account and which results in a distinction between single and composite indicators.
- A classification that takes into account at what point in the project life cycle we apply indicators which results in a distinction between performance indicators, outcome and impact indicators.

The first two classifications are attributed to the social indicator movement, while the third classification links with programme evaluation. I expand on the three classifications below.

Subjective vs objective (Also referred to as qualitative vs quantitative)

Early social indicators were originally categorised as objective indicators due to the strong influence of the empirical tradition and the work of early pioneers such as William Ogburn (Callebaut, 1978; Carley 1981, Diener and Suh, 1997, Rossi & Gilmartin, 1980). Ogburn placed significant emphasis on the objectivity of social indicator reporting and the need for social reports to be bias and value free. A major contribution of the social indicator movement was the increasing interest and advancements in subjective indicators brought to the fore by methodological debates surrounding the measurement of quality of life (Andrews, 1974, Andrews and Withey, 1975, Diener, 2005, Noll, 2002b).

Rossi and Gilmartin (1980, p.19) posit that “subjective indicators are based on the reports persons make about their feelings, attitudes, and evaluations. Objective indicators on the other hand are based on counts of behaviours and conditions associated with given situations”. Carley (1981, p.31) in his

extensive discussion of objective vs subjective indicators, introduces the different levels at which these two types of indicators are measured:

> Objective indicators are the occurrences of given phenomena, such as environmental stimuli and behavorial responses, which are measurable on an interval or ratio scale, and amenable to the usual methods of data analysis…Subjective indicators on the other hand, are those based on reports from individuals on the "meaning" aspects of their reality and as such represent psychological variables which are usually presented on an ordinal scale. Questionnaires, interviews and opinion polls elicit this subjective information.

Ratio measurement has a true zero point (for example age) and can therefore be subjected to all types of analytics (average, mean, medium etc.). With interval measurement, the distance between the variables have meaning –i.e. the difference between 80 degrees and 90 degrees Celsius are the same as the distance between 40 and 50 degrees Celsius. However, 160 degrees Celsius is not twice as hot as 80 degrees Celsius given that there is no true zero point (as there is not a complete lack of heat). With interval data, one is therefore able to undertake some level of analysis but less so than with ratio level data which has a true zero point. Nominal and ordinal data are "categories of things that can be counted", and that can be treated accordingly (Gorard, 2010, p.395). These categories therefore reflect the attributes of the variables we are measuring and are therefore not a "true" number as would be the case with ratio and interval level measurement (Babbie and Mouton, 2001). Nominal and ordinal data is distinguished on the basis that the latter contains an "intrinsic order" (Gorard, 2010, p.396). Compared to ratio and interval measurement and analysis, nominal and ordinal data offer the least amount of analytical capabilities.

Figure 24 shows that the nominal or categorical level of measurement is the weakest as differences between variables are only named. Ordinal level measurement introduces differences in ordering and rank. At the interval level (e.g. in intelligence testing or temperature) the distances between measures become meaningful. But the highest level – and the most useful – is the ratio level measurement because of the absolute zero point. At this level we speak of 'real numbers' and how to measure the intervals between them as true quantities, such as differences in age, income, etc.

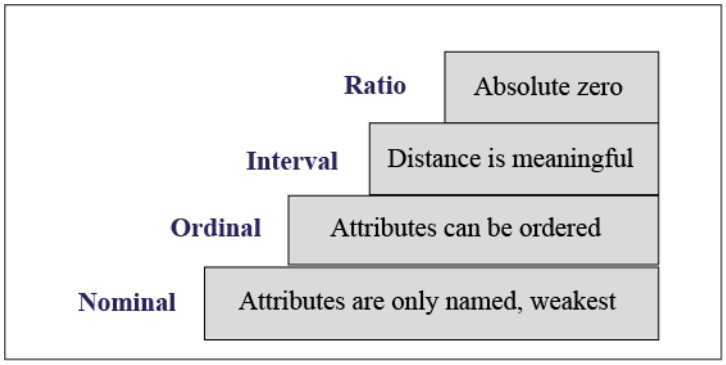


Figure 24: Levels of measurement

The distinction between 'quantitative' and 'qualitative' indicators derives from the underlying level at which we measure objects or phenomena. When our measurements produce true numbers (number of people infected, proportion of people in an age interval studying at universities, rate of growth in research output by year, number of alcohol-related accidents per week, etc.) these measures are then typically used as quantitative indicators of some underlying constructs such as (infection rate, education participation rates, research productivity, extent of alcohol use).

When our measurements simply generate a number that is a 'placeholder' for a category (yes/no responses) or a label for a rank-ordering of responses to a scale of some kind, these measures are defined as qualitative indicators of some underlying construct (the presence or absence of something or extent of satisfaction to service delivered).

- Nominal measures describe variables with attributes that are simply different from one another, such as "gender".
- Ordinal measures refer to phenomena with attributes we can rank-order along some progression from more to less. One example is the phenomenon "prejudice" as composed of the attributes very prejudiced, somewhat prejudiced, slightly prejudiced, and not at all prejudiced.
- Interval measures refer to those constructs with attributes not only rank-ordered but also separated by a uniform distance. One example is IQ.
- Ratio measures are the same as interval measures except that ratio measures are also based on a true zero point. Age is an example of a ratio measure, since that variable contains the attribute zero years old.

A given variable or construct can sometimes be measured at different levels of measurement. Thus, age, potentially a ratio measure, may also be treated as interval, ordinal, or even nominal. The most appropriate level of measurement used depends on the purpose of the measurement.

Single versus composite indicators

Another methodological contribution of the social indicator movement were the discussions around the aggregation of indicators. The complexities of combining social indicators into a singular value has been a topic that features prominently across the social indicator literature (Craig and Driver, 1972 as cited in Carley, 1981).

This classification principle captures the *complexity* of the construct or phenomenon being measured. The "quality of life" work stream of the social indicator movement triggered discussion about the best way to measure 'quality of life' as it was recognised that a combination of indicators is needed to capture this construct's multi-dimensional nature. When combining indicators, invariably issues surrounding the underlying weighting of each indicator comes to the fore. Even though the social indicator movement did not resolve these challenges; it did advance the thinking around how to measure complex phenomena. Composite indicators entail the development of dimensions, and should not be confused with an aggregate set of indicators:

- Aggregate set of indictors: a set of single (simple) indicators are aggregated (weighted or unweighted) into a single value. No development of dimensions take place.
- Composite indicators: A composite indicator is formed when individual indicators are compiled into a single index, on the basis of an underlying model of the multi-dimensional concept that is being measured (OECD, 2004).

To summarise: The phenomena or objects that we wish to measure (the "constructs") can range from reasonably simple to very complex. This range also coincides with the levels of analysis. Typically, at the level of intervention programmes, we target individuals (the intended beneficiaries) and aim to change some aspect of their behaviour, attitudes, beliefs, skills, competencies, etc. At the level of institutions, we typically want to measure and evaluate some more complex features such as 'business performance', 'firm profitability', 'organisational learning', 'hospital management efficiency', etc. And at the systems level (e.g. the science system) we typically are interested in very high-level complex properties of the system: the transformation of our educational system, the functionality of our schools,

etc. The more complex the object is that we wish to measure, the more it becomes necessary that we move from single or individual indicators to composite indicators and indices.

Classification of indicators in terms of the intervention life cycle

The third way of classifying indicators of relevance to this study, pertains specifically to the application of indicators when we conduct performance monitoring and evaluations. Programme evaluation subscribes to a system theory in that "it must ensure the smooth transition of inputs into desirable outputs" with outputs donating both outputs *and* outcomes (Chen, 2005, p.5).

The systems approach to social programmes is carried through in the structure of the logic model – with indicators then developed for each of the system elements: i.e. input indicators, output indicators, outcome indicators and impact indicators.

The above clarification of key social measurement concepts and indicator categories guided the development of the eLearning Game Changer indicators. The following will be covered in more detail in the next section:

- Identifying the constructs in the outcome statement and operationalising these constructs in order to make them more measurable.
- Describing the indicators used to measure the overarching construct of eLearning, utilising different categories of indicators, namely:
    - Quantitative vs qualitative indicators
    - Single vs aggregate vs composite indicators
    - Intervention level indicators, specifically performance indicators vs outcome indicators

8.4.3 Developing the indicators for the eLearning Game Changer

This section will draw on the eLearning data plan where I distinguish between performance indicators (Annexure B1) and outcome indicators (Annexure B2). I first discuss performance indicators before providing a detailed reflection on the outcome indicators.

Performance indicators are an indicator type used in programme evaluation, but which has its origin in the performance measurement tradition (Ref. section 2.3.1 under the NPM). An indicator becomes a performance indicator when a target is attached to the indicator. In Chapter 2 I discussed the meaning

of performance according to the 3 Es – economy, effectiveness and efficiency, positing that performance indicators should not only stop with the measurement of inputs and outputs but also outcomes. However, without a target this technically would not constitute a performance indicator but merely an outcome indicator. Regardless of these technicalities, when an indicator has a target attached to it, it can be categorised as a performance indicator (McLaughlin & Jordan, 2010).

Many of the eLearning Game Changer outputs had targets attached, and therefore these indicators were referred to as performance indicators and not output indicators. The table below shows some examples of the performance indicators included in Annexure B1 of the data plan:

Table 21: Examples of performance indicators

Outputs	Performance indicator
1239 schools connected to broadband	Number of schools connected to broadband (Target = 1239)
170 Enhanced schools with LANs	Number of Enhanced schools with LANs (Target =170)
16 Model schools with learner management systems	Number of model schools with learner management systems implemented (Target=16)

The only exception was the indicator attached to measuring the ICT and technology related queries logged at the CeI service desk[13]: Number of queries logged at CeI service desk per category of query. This indicator has no target, given that there was no means in predicting the number of queries that were logged at the service desk. This indicator is therefore classified as an output indicator and not a performance indicator.

I now move to outcome indicators. The starting point in formulating outcome indicators was the identification of the relevant construct to be measured (conceptualisation step). Some examples of the constructs contained in the outcome statements of the eLearning Game Changer are shown as underlined text below:

- Principals support the eLearning Game Changer
- Schools in three categories have better connectivity

[13] The CeI handles all ICT related aspects in the province. A service desk was established where all queries are logged for resolution, including all schools in the province

- Targeted teachers integrate technology and digital resources into their teaching environment
- Curriculum support officials provide better ICT support to teachers

Each outcome statement should ideally only contain one construct to ensure each construct is addressed separately in terms of measurement. In order to ensure conceptual clarity about what needs to be measured questions were asked such as "What would 'principal support' for the Game Changer look like?", "What is meant by access?" and "How will we know that teachers are integrating ICT?". These questions served the purpose of getting us one step closer to understanding these abstract notions and developing indicators to measure them.

An example is provided below:

Outcome statement: Principals support the eLearning Game Changer

Operational definition: Number of school principals in public ordinary schools that show support to the eLearning as indicated by the fact that they have policies, plans and people in place to implement eLearning in the school

Indicator: Number of school principals that express commitment to eLearning Game Changer

Indicator calculation (technical description)**:** count if indicated "yes" to questions 1-7 of annual school survey. Average commitment to eLearning: 31-60% of yes responses versus a good commitment to eLearning (61% and higher). Express the number of principals with a score of 61% and higher as a percentage of those principals that answered this question in the annual school survey. Disaggregate at the level of school category: model, enhanced and universal

Even though the indicator (Number of school principals that express commitment to eLearning Game Changer) did not necessarily contain a fully operationalised description of the construct (i.e. commitment), attempts were made to provide operational definitions to as many of these statements as possible in the eLearning data plan.

Drawing on the three indicator categories introduced above, the mix of indicators employed in the eLearning Game Changer is discussed further below.

8.4.4 Applying the three indicator classifications to the eLearning Game Changer

Table 22 contains a categorisation of the original 45 eLearning Game Changer indicators along the three categories described above: i) single vs composite vs aggregate, ii) quantitative vs qualitative and iii) output vs performance vs outcome vs impact indicators. This categorisation demonstrates the diverse mix of indicators needed to measure a complex construct such as eLearning. Some examples are pulled from Table 22 as the three categories are discussed.

The complexity of the phenomenon (or construct) being measured determines whether a single indicator will suffice or whether a more sophisticated way of measurement is needed. In the case of the eLearning Game Changer a mix of single, aggregate, and composite indicators (definitions provided to the left), were needed to measure the outcomes and outputs.

Recapping the terminology…

Single indicator: one indicator providing a single value which captures the presence or absence of the phenomenon under study

Aggregate indicator: a set of indicators are aggregated or summed (weighted or unweighted) into a single value. No development of dimensions take place

Composite indicator: A composite indicator is formed when individual indicators are compiled into a single index, on the basis of an underlying model of the multi-dimensional nature of the concept that is being measured (OECD, 2004)

All the performance indicators and output indicator are single indicators, i.e. only one indicator is needed to measure the accompanying output. Given the abstract nature of outcomes, a mix of indicators were needed which covered single indicators, a set of single indicators, aggregated indicators, and composite indicators.

I will start with an example where a single indicator was sufficient to measure the outcome. The quality of technical support provided to schools was measured by one indicator, viz. the percentage of queries reported to CeI service desk that are resolved within five working days.

Outcome statement: Schools receive better technological support

Single indicator: Number of queries reported to CeI service desk that are resolved within five working days

For some of the outcomes, a single indicator was not deemed sufficient so then multiple indicators were constructed. The example below shows how the notion of "the integration of digital resources by teachers" was measured by three indicators:

Outcome statement: Teachers integrate digital resources into their teaching environment

Indicator (1): Number of teachers that access professional learning communities on e-Portal and through other collaborative platforms

Indicator (2): Number of teachers that upload content onto e-Portal

Indicator (3): Number of teachers that share resources with other teachers *within* and *outside* of school

The extract from the Teacher survey below shows how these three indicators were defined in the questionnaire. Indicator 3 was covered by two questions in the teacher survey (E7 and E8 below) and combined to reflect teachers who indicated that they either shared resources within their own school and/or outside. Indicators 1 and 2 were obtained from question E9 (See option 1 and 3):

Extract from Teacher survey

E7. Do you share your digital resources with other teachers in your school? (tick one)

Yes – I share my resources	No – I don't share my resources	Not applicable - I don't develop my own digital content

E8. Do you share your digital resources with teachers outside of your school? (tick one)

Yes – I share my resources	No – I don't share my resources	Not applicable - I don't develop my own digital content

IF TEACHER SELECTED YES TO E8, CONTINUE TO E9, ELSE SKIP TO SECTION F

E9. How do you share your digital resources with teachers outside of your school?

I place it on the ePortal	I share it via email/ WhatsApp/ Facebook etc.	I share it informally with other teachers/ via the professional learning community	Not applicable - I don't develop my own digital content	Other, please specify

An example of where the indicators were combined into an aggregate indicator relates to the outcome statement: Principals support the eLearning Game Changer.

Outcome statement: Principals support the eLearning Game Changer

Aggregate outcome indicator: Number of principals that support the eLearning Game Changer

Operational definition: Number of school principals in public ordinary schools that show good support to the eLearning as indicated by the fact that they have policies, plans and people in place to implement eLearning in the school

The underlined text in the operational definition represents seven questions in the principal survey as shown in the table extracted from the survey:

Extract from Principal survey

Please indicate the status of the following with regards to ICT at your school? *(Tick one option per statement*

	Yes	No	Not sure	Busy developing
1. Our school has a policy on allocation of technology-enabled classrooms to teachers/ for specific subjects (e.g. smart classrooms)				
2. Our school has an ICT committee				
3. Our school has an ICT champion				
4. Our school has a cell phone policy for learners that allows access to the internet / use during school hours for learning purposes				
5. Our school has a cell phone policy for teachers that allows access to the internet / use during school hours for teaching purposes				
6. Our 2018 school improvement plan includes the use of ICT to improve language and maths				
7. We have a teacher development plan in place that provides for professional development and support for maths & language teachers to integrate technology into pedagogy				

The indicator calculation demonstrates how the responses from the 7 questions were summed to arrive at a single value for “principal support” – see indicator calculation below:

Indicator calculation: count if indicated "yes" to questions 1-7 and express as percentage the principals that show average support for eLearning (31-60% of yes responses) versus good support for eLearning (61% and higher).

No dimensions were however developed; each question represent the attributes associated with the variable, namely: "support". Because we needed a single "value" for support the seven questions' responses were weighted equally and calculated as described above.

A composite indicator was developed for the eLearning Game Changer for the outcome "targeted teachers integrate available technology into their teaching environment". In order to derive at a more measurable version of "ICT Integration", four dimensions were identified: using ICT to communicate, using ICT in the classroom, using ICT for learning management purposes and using ICT for learner management. The operational definition reflects these four dimensions:

Outcome statement: Teachers integrate available technology into their teaching environment

Composite outcome indicator: Number of targeted teachers that integrate available technology into their teaching environment

Operational definition: Number of teachers that integrate ICT in a) their communication with parents, fellow teachers and learners, ii) that use ICT in the classroom, iii) that use ICT for learning management systems and iv) that use ICT in learner management systems

Each of the dimensions was measured by between 1 and 4 questions in the teacher survey, asking teachers to indicate what they use ICT for.

Extract from Teacher survey (Use of ICT)

D1. Please indicate in which areas you use ICT currently? (tick all relevant options)	I don't use ICT at all	
	I send e-mails/sms/WhatsApp's or use other electronic methods to communicate with learners around academic and school matters	
	I send e-mails/sms/WhatsApp's or use other electronic methods to communicate with parents around academic and school matters	
	To communicate with fellow teachers around academic and school matters	
	Internal administration: inputting on documents, timetables, etc.	
	I use a learner management system (daily recording of attendance registers, learner profiles and learner scores electronically)	
	I use a learning management system to upload digital content or link to digital content	
	I use a learning management system for lesson planning (e.g. Moodle, MS Classroom or Google Classroom)	

	I use a learning management system to deliver digital content in the classroom (e.g. Moodle, MS classroom or google Classroom)	
	Teaching and Learning in classrooms	
	Conducting online assessments	
	Other, please specify	

The text box below demonstrates the indicator calculation as well as weighting of the four dimensions, as well as the survey questions linked to the four dimensions:

Dimension 1: Using ICT in classroom: 60%
Questionnaire item: using ICT for Classroom teaching
Questionnaire item: conducing online assessments

Dimension 2: Using ICT to communicate: 5%
Questionnaire item: using ICT to communicate with teachers
Questionnaire item: using ICT to communicate with learners
Questionnaire item: using ICT to communicate with parents

Dimension 3: Using ICT for learning management: 20%
Questionnaire item: using ICT for learning management (uploading digital content)
Question item: using a learning management system (to deliver digital content in the classroom)
Questionnaire item: using ICT for lesson planning

Dimension 4: Using ICT for admin: 15%
Questionnaire item using ICT for internal administration
Questionnaire item: using ICT for learner management

Figure 25: Weighted breakdown of ICT integration dimensions

The distinguishing factor between an aggregated indicator and composite indicator is the assignment of dimensions. In the ICT Integration example, four dimensions were identified; whereas in the principal support example no dimensions were identified; the scores on individual items were simply summed or aggregated to derive at a single value of principal support.

Both objective (quantitative) and subjective (qualitative) indicators have a role to play in the measurement of social phenomena. Whereas quantitative indicators provide an objective view of programme roll out and progress, often the “hard numbers” fall short in conveying the perspectives of the programme beneficiaries. Qualitative indicators on the other hand capture beneficiaries’ attitudes or perceptions about a particular topic or intervention; something which quantitative data are unable to convey. A substantial number of eLearning outcome indicators can be categorised as qualitative due to the fact that surveys (i.e. self-reporting data) constituted a large part of the data collection activities.

Such self-reported data provided teachers, learners, and principals the opportunity to express their perceptions and experiences on access to, attitudes towards and their use of ICT. Examples of qualitative indicators include:

- Number of school principals that support for the eLearning Game Changer.
- Number of teachers at model schools that use learning management systems to plan and deliver online lessons.
- Number of principals that are aware of the ICT Integration support available to them.
- Number of model and targeted enhanced school principals that report improved data collection and management (school admin system and data dashboard).

A full categorisation of the eLearning indicators, covering all three categories discussed above, are shown in the table below. The performance and output indicators are indicated with grey highlights while the impact indicators are listed at the bottom of the table.

Table 22: Classification of eLearning indicators per work stream

Work stream	Outcomes/ Output statements	Indicators	Indicator categorisation		
			Single/ Aggregate/ Composite	Quantitative/ Qualitative	Indicator type
eInfrastructure	1278 schools have WAN	Number of schools connected to Broadband (per school category)	Single	Quantitative	Performance indicator
	435 enhanced schools have LAN	Number of enhanced schools with LANs	Single	Quantitative	Performance indicator
	972 Schools with wireless access points	Number of schools with wireless access points as part of the slim lab rollout (universal and enhanced schools)	Single	Quantitative	Performance indicator
	School ICT queries tracked (WCG and non-WCG related)	Number of queries logged at CeI service desk (per category of query)	Single	Quantitative	Output indicator
	Teachers have better access to internet connectivity	Number of teachers that connect to wireless access points to enable own device (per school category)	Single	Quantitative	Outcome indicator
	Schools have better access to internet connectivity	Data being accessed and downloaded by schools (by type, per school category)	Single	Quantitative	Outcome indicator
		Schools' data volumes (per school category)	Single	Quantitative	Outcome indicator
	Schools receive better technological support	Number of queries reported to CeI service desk that are resolved within five working days (per category of query, per school category)	Single	Quantitative	Outcome indicator
eTechnology	1160 schools have slim laboratory refreshed	Number of schools with slim laboratory refresh (universal and enhanced schools)	Single	Quantitative	Performance indicator
	7530 smart classrooms deployed with teacher devices	Number of smart classrooms (teacher devices) deployed (model and enhanced schools)	Single	Quantitative	Performance indicator

Work stream	Outcomes/ Output statements	Indicators	Indicator categorisation		
			Single/ Aggregate/ Composite	Quantitative/ Qualitative	Indicator type
	16183 devices distributed to model schools	Number of learner devices that has been distributed to model schools	Single	Quantitative	Performance indicator
	Learners have improved access to technology	Percentage of maths & language contact time spent in smart classroom environment (For mathematics & language, per grade, per school category)	Single	Quantitative	Outcome indicator
		Percentage of maths & language contact time spent in laboratory environment (For mathematics & language, per grade, per school category)	Single	Quantitative	Outcome indicator
		Number of model school learners that connect devices on a daily basis	Single	Quantitative	Outcome indicator
eTeachers/ eOfficials	32214 Principals and teachers trained in basic ICT competencies	Number of school staff trained in basic ICT competencies (principals and teachers; per school category)	Single	Quantitative	Performance indicator
	4056 teachers trained in ICT integration	Number of teachers trained in ICT integration (per school category)	Single	Quantitative	Performance indicator
	1239 Principals and school management officials trained in ICT integration	Number of school management officials trained (Principals, deputy principals, HoDs; per school category)	Single	Quantitative	Performance indicator
	Teachers are aware of the ICT integration support system in place at districts	Number of teachers that are aware of the ICT Integration support available to them (per school category)	Single	Qualitative	Outcome indicator
	Principals are aware of the ICT integration support system in place at districts	Number of principals that are aware of the ICT Integration support available to them (per school category)	Single	Qualitative	Outcome indicator

Work stream	Outcomes/ Output statements	Indicators	Indicator categorisation		
			Single/ Aggregate/ Composite	Quantitative/ Qualitative	Indicator type
	Principals receive ICT integration support from district	Number of principals that request support from district staff (per school category)	Single	Qualitative	Outcome indicator
	Teachers receive ICT integration support from district	Number of teachers that request support from District staff (per school category)	Single	Qualitative	Outcome indicator
	Curriculum support officials provide better ICT support to principals	Number of principals that are satisfied with support they receive from the district (per school category)	Single	Qualitative	Outcome indicator
	Curriculum support officials provide better ICT support to teachers	Number of teachers that are satisfied with support they receive from the district (per school category)	Single	Qualitative	Outcome indicator
	Teachers use appropriate technology in their teaching practice	Number of teachers that use wireless access points to enrich educational practices (per school category)	Single	Quantitative	Outcome indicator
		Number of teachers at model schools that use learning management systems to plan and deliver online lessons	Single	Qualitative	Outcome indicator
	Teachers integrate appropriate and available technology into their teaching environment	Number of teachers that integrate ICT into their teaching environment (per school category)	Composite	Qualitative	Outcome indicator
eContent	Five digital resources available for every Maths and Language topic grade 4-12	Number of topics with digital resources available (per subject, per grade)	Single	Quantitative	Performance indicator

Work stream	Outcomes/ Output statements	Indicators	Indicator categorisation		
			Single/ Aggregate/ Composite	Quantitative/ Qualitative	Indicator type
	Learners have better access to digital resources	Number of learners that indicate that a) they choose digital resources as a first choice when learning and b) they find digital resources stimulating and useful (per school category)	Single	Qualitative	Outcome indicator
		Count of CAPS aligned downloaded/ page views by learners (per subject, per grade, per school category)	Single	Quantitative	Outcome indicator
	Teachers integrate digital resources into their teaching environment	Number of teachers that access professional learning communities on e-Portal and through other collaborative platforms on a recurring basis (per school category)	Single	Quantitative	Outcome indicator
		Number of teachers that upload content onto e-Portal (per school category)	Single	Qualitative	Outcome indicator
		Number of teachers that share resources with other teachers through social media and electronic communication (per school category)	Single	Qualitative	Outcome indicator
eAdmin	16 model schools, 50 enhanced schools, 25 universal schools have web-based school admin systems	Number of schools with web-based school administration system in place (per school category)	Single	Quantitative	Performance indicator
	16 model schools, 150 enhanced schools, have data dashboards	Number of schools where data dashboards have been made available/ rolled out (model and enhanced)	Single	Quantitative	Performance indicator

Work stream	Outcomes/ Output statements	Indicators	Indicator categorisation		
			Single/ Aggregate/ Composite	Quantitative/ Qualitative	Indicator type
	16 model schools, 10 enhanced schools have learning management systems	Number of schools with learning management system implemented (model and enhanced)	Single	Quantitative	Performance indicator
	16 Model schools, 10 Enhanced schools have parent/learner portal	Number of schools with parent/ learner portal implemented (model and enhanced)	Single	Quantitative	Performance indicator
	Targeted principals have better access to digital school admin systems and school dashboards	Number of schools that upload learner assessment results to a central repository (model and targeted enhanced)	Single	Quantitative	Outcome indicator
		Number of schools that access learner assessment and attendance data on the dashboard (model and targeted enhanced schools)	Single	Quantitative	Outcome indicator
	Targeted principals use the digital school admin systems and school dashboards	Number of schools that use a school admin system to collect school based assessment and attendance data (model and targeted enhanced)	Single	Quantitative	Outcome indicator
		Number of school principals that report improved data collection and management (model and enhanced)	Single	Qualitative	Outcome indicator
eCulture	1499 schools sign Memorandum of understanding (MOU)/ memorandum of agreement (MOA) or Letter of commitment	Number of schools that sign MOU/MOA/ Letter of commitment (per school category)	Single	Quantitative	Performance indicator
	School principals support the eLearning GC	Number of school principals that express commitment to eLearning Game Changer (per school category)	Aggregate	Qualitative	Outcome indicator

Work stream	Outcomes/ Output statements	Indicators	Indicator categorisation		
			Single/ Aggregate/ Composite	Quantitative/ Qualitative	Indicator type
Learner Impact Level	Enhanced motivation to learn	Number of learners attending class (per Grade, per school category)	Single	Quantitative	Impact indicator
	Improved learner outcomes	Learner performance scores (mathematics & language; per grade, per school category)	Single	Quantitative	Impact indicator
	Increased classroom participation	Number of learners who actively participate in classroom activities (per selected grades)	Aggregate	Qualitative	Impact indicator

8.5 Application of the eLearning Game Changer's indicators

In Chapter 5 I provided an overview of the methodological choices for the eLearning Game Changer. To recap: both sampling and selection of cases were undertaken for new data collection, whereas with existing data, all available data were utilised. In terms of data collection methods, a combination of unobtrusive and obtrusive methods of data collection were used. All the methods (unobtrusive and obtrusive) were quantitative in nature.

The table below, presented in Chapter 5, overlays the sampling and selection choices with the data collection methods as well as data collection tools (distinguished between reactive and non-reactive).

Table 23: Data collection methods (and tools): split between reactive and non-reactive

Data collection methods and tools		New/ Existing/ Expanded	Sampling/selection of schools (cases)
REACTIVE (OBTRUSIVE)	**Surveys**		
	Teacher and learner questionnaires	New data	63 schools
	Principal questionnaire	Customised data	All schools
	Timetable schedule	New data	47 schools
	Observations		
	Classroom observation schedule	New data	16 schools
NON-REACTIVE (UNOBTRUSIVE)	**Platform and implementation data**		
	Learner and teacher level WAN usage report	New data	111 schools
	WAN downtime report	New data	111 schools
	eCulture checklist	New data	All schools
	CEMIS system report	Customised data	All schools
	Technology roll out checklists	Existing data	All schools
	CEI Service desk report	Customised data	All schools
	ePortal usage report	Customised data	All schools
	Learner and teacher data		
	Teacher training report	Existing data	All schools
	Learner recording and reporting sheets	Existing data	111 schools

I now show the link between the data collection tool and the indicators.

Table 24: Data collection tool per indicator

Work stream	Output and Outcome statements	Performance and Outcome Indicators	Data collection tool
eInfrastructure	1278 schools have WAN	Number of schools connected to Broadband (per school category)	Technology roll out checklists
	366 enhanced schools have LAN	Number of enhanced schools with LANs	Technology roll out checklists
	972 Schools with wireless access points	Number of schools with wireless access points as part of the slim lab rollout (universal and enhanced)	Technology roll out checklists
	School ICT queries tracked	Number of queries logged at CeI service desk per category of query	CeI Service desk report
	Teachers have better connectivity	Number of teachers that connect to wireless access points to enable own device (per school category)	Learner and teacher level WAN usage report
	Schools have better connectivity	Data being accessed and downloaded by schools (by type, per school category)	WAN usage report
		Schools' individual data volumes (per school category)	WAN usage report
	Schools receive better technological support	Number of queries reported to CeI service desk that are resolved within five working days (per category of query; per school category)	CeI Service desk report
eTechnology	1160 schools have slim laboratory refreshed	Number of schools with slim laboratory refresh (universal and enhanced)	Technology roll out checklists
	7530 smart classrooms deployed with teacher devices	Number of smart classrooms (teacher devices) deployed (model and enhanced)	Technology roll out checklists
	16183 devices distributed to model schools	Number of learner devices that has been distributed to model schools	Technology roll out checklist
	Learners have better access to technology	Percentage maths & language contact time spent in smart classroom environment (For mathematics & language, per grade, per school category)	Timetable schedule
		Percentage maths & language contact time spent in laboratory environment (For mathematics & language, per grade, per school category)	Timetable schedule
		Number of model school learners that daily connect devices	Learner and teacher level WAN usage report

Work stream	Output and Outcome statements	Performance and Outcome Indicators	Data collection tool
eTeachers/ eOfficials	1239 Principals and school management officials trained	Number of school management officials trained (Principals, deputy principals, HoDs; per school category)	Teacher training report
	32214 Principals and teachers trained in basic ICT competencies	Number of school staff trained in basic ICT competencies (principals and teachers; per school category)	Teacher training report
	4056 teachers trained in ICT integration	Number of teachers trained in ICT integration (per school category)	Teacher training report
	Teachers are aware of the ICT integration support system in place at districts	Number of teachers that are aware of the ICT integration support available to them (per school category)	Teacher questionnaire
	Principals are aware of the ICT integration support system in place at districts	Number of principals that are aware of the ICT integration support available to them (per school category)	Principal questionnaire
	Principals receive ICT integration support from district	Number of principals that request support from district staff (per school category)	Principal questionnaire
	Teachers receive ICT integration support from district	Number of teachers that request support from district staff (per school category)	Teacher questionnaire
	Curriculum support officials provide better ICT support to principals	Number of principals that are satisfied with support they receive from the district (per school category)	Principal questionnaire
	Curriculum support officials provide better ICT support to teachers	Number of teachers that are satisfied with support they receive from the district (per school category)	Teacher questionnaire
	Teachers use available technology in their teaching practice	Number of teachers that use wireless access points to enrich educational practices (per school category)	Learner and teacher level WAN usage report
		Number of teachers at model schools that use learning management systems to plan and deliver online lessons	Teacher questionnaire

Work stream	Output and Outcome statements	Performance and Outcome Indicators	Data collection tool
	Teachers integrate available technology into their teaching environment	Number of teachers that integrate ICT into their teaching environment (per school category)	Teacher questionnaire
eContent	Five digital resources available for every Maths and Language topic grade 4-12	Number of topics with digital resources available (per subject; per grade)	ePortal usage report
	Learners have better access to digital resources	Number of learners that indicate they a) choose digital resources as a first choice when learning and b) find digital resources stimulating and useful (per school category)	Learner questionnaire
		Count of CAPS aligned downloaded/ page views by learners (per subject, per grade, per school category)	ePortal usage report
	Teachers integrate digital resources into their teaching environment	Number of teachers that access professional learning communities on e-Portal and through other collaborative platforms on a recurring basis (per school category)	Teacher questionnaire
		Number of teachers that upload content onto e-Portal (per school category)	Teacher questionnaire
		Number of teachers that share resources with other teachers through social media and electronic communication (per school category)	Teacher questionnaire
eAdmin	16 model schools, 50 enhanced schools, 25 universal schools have web-based school admin systems	Number of schools with web-based school administration system in place (per school category)	Technology roll out checklist
	16 model schools, 150 enhanced schools, have data dashboards	Number of schools with access to data dashboards (model and enhanced)	Technology roll out checklist
	16 model schools, 10 enhanced schools have learning management systems	Number of schools with learning management system implemented (model and enhanced)	Technology roll out checklist
	16 Model schools, 10 Enhanced schools have parent/learner portal	Number of schools with parent/ learner portal implemented (model and enhanced)	Technology roll out checklist

Work stream	Output and Outcome statements	Performance and Outcome Indicators	Data collection tool
	Principals have better access to digital school admin systems and school dashboards	Number of schools that upload learner assessment results to a central repository (model and targeted enhanced)	CEMIS system report
		Number of schools that access learner assessment and attendance data on the dashboard (model and targeted enhanced schools)	CEMIS system report
	Principals use the digital school admin systems and school dashboards	Number of schools that use a school admin system to collect school based assessment and attendance data (model and targeted enhanced)	CEMIS system report
		Number of school principals that report improved data collection and management (model and enhanced)	Principal questionnaire
eCulture	1499 schools sign MOU or MOA	Number of schools that sign MOU/MOA/ Letter of commitment (per school category)	eCulture checklist
	School principals support the eLearning Game Changer	Number of school principals that express commitment to eLearning Game Changer (per school category)	Principal questionnaire
Learner Impact Level	Enhanced motivation to learn	Proportion of learners attending class (per Grade, per school category)	Learner recording and reporting sheets
	Improved learner outcomes	Learner performance scores (mathematics & language; per grade, per school category)	Learner recording and reporting sheets
	Increased classroom participation	Percentage of learners who actively participate in classroom activities (per selected grades)	Classroom observation schedule

In February 2017, a list of the indicators to be tracked per work stream and the accompanying data collection tools was presented to stocktake for approval. Once approved, the DSU set out to develop a detailed data plan that would assign roles and responsibilities for the data collection functions. The data plan (Annexure B1 and B2) shows the organisation (s) responsible for collecting or obtaining the data. As performance tracking director in the DSU my role was to provide oversight of the entire data plan ensuring all data was provided as stipulated.

Table 25 provides a breakdown of the data collection responsibilities between the DSU, CeI and WCED as set out in the data plan. Responsibility in this instance means the organisation or unit responsible for ensuring the data is collected and/or obtained, not necessarily undertaking the data collection. This is an important distinction. For example, the DSU had to ensure the learner, teacher and principal questionnaires were completed as planned and therefore put the needed systems in place to enable this; however the actual collection of data was undertaken with the assistance of the eLearning advisors in the eight educational districts. Similarly, WCED was responsible for obtaining the learner recording and reporting sheets from the sample schools; but it is the schools that collected the data by completing the learner reporting sheets.

Table 25: Organisations responsible for data collection

Data collection methods and tools		Sampling/selection of schools (cases)	New/ Existing	Organisation responsible
REACTIVE (OBTRUSIVE)	**Surveys**			
	Teacher and learner questionnaires	63 schools	New data	DSU
	Principal questionnaire	All schools	Customised data	DSU
	Timetable schedule	47 schools	New data	WCED
	Observations			
	Classroom observation schedule	16 schools	New data	DSU (with WCED)
NON-REACTIVE (UNOBTRUSIVE)	**Platform and implementation data**			
	Learner and teacher level WAN usage report	111 schools	New data	CeI
	WAN downtime report	111 schools	New data	CeI
	eCulture checklist	All schools	New data	WCED
	CEMIS system report	All schools	Customised data	WCED

Data collection methods and tools		Sampling/selection of schools (cases)	New/ Existing	Organisation responsible
	Technology roll out checklists	All schools	Existing data	WCED
	CEI Service desk report	All schools	Customised data	CeI
	ePortal usage report	All schools	Customised data	WCED
	Learner and teacher data			
	Teacher training report	All schools	Existing data	WCED
	Learner recording and reporting sheets	111 schools	Existing data	WCED

More extensive arrangements needed to be put in place for collecting new data as opposed to working with existing data sources. This was particularly pertinent as far as data collection from the 111 sample schools was concerned (which includes the 63 eLearning Game Changer schools).

The WCED, in consultation with the district directors, tasked the eLearning advisors to assist with the data collection activities in the 111 schools. This included distributing the teacher and learner questionnaires, completing the timetable schedule and obtaining the learner recording and reporting sheets from the sample schools. The DSU convened various engagements with district staff and eLearning advisors over the two year period to explain the purpose of the Game Changer data activities, to discuss and describe the various data collection activities and to equip them with the skills needed to collect the required data.

Getting the 111 sample schools on board was vitally important. The first engagement with the 111 sample schools took place during June/ July 2017 whereby the DSU and officials from WCED visited the 8 educational districts to convene an information session with the 111 schools. These sessions explained the purpose of the Game Changers and the importance of data in demonstrating results, requesting the 111 schools to come on board for a two-year period. Most schools committed to provide data over the two years and consent forms were signed by the principals. These consent forms provided the DSU with permission to collect learner data for a period of two years.

The 111 schools were asked to identify a data coordinator in their school that the specific eLearning advisor could liaise with. In some instances, the principal opted to be the data coordinator but in many instances a teacher or other staff member was assigned the responsibility. This staff member or teacher

would be the person the eLearning advisor would liaise with every time data collection was undertaken. Formalising the process in this way paid huge dividends: the eLearning advisors had one person they could liaise with whenever the data collection activities took place, and from the DSU and WCED head office's perspective they had feet on the ground to ensure the data was collected.

8.6 Reflections on the eLearning performance measurement

The biggest challenge related to performance tracking in the Game Changers was the transition to a truly evidence-based government. The WCG was largely "evidence-based in name only". The lack of data availability and budget to collect the required data coupled with the limited departmental data expertise meant the DSU had to play a much bigger role in the performance measurement activities than originally envisaged. This is not to argue that WCED and the broader WCG were not investing in their data systems – at the time of the Game Changers, a programme was underway in WCG to enhance the use of data but this would take time to implement, and would not meet the specific eLearning Game Changer data requirements.

The following were some of the very real problems and obstacles that needed to be addressed in this regard:

- Developing outcomes and outcome indicators were challenging for programme staff. It required many facilitated sessions to derive at the outcomes for the Game Changers. Conceptually, programme staff battled with the notion of outcomes, and even more so with the concept of indicators.
- The availability of, and quality of existing data was over estimated: the WCED collected extensive data from schools but this was not done at teacher and learner level. Data was typically provided in aggregated format which limited its usefulness. This required that additional tools be developed to ensure that all data needs are addressed, as well as putting new systems in place to ensure the needed data is collected.
- The time taken to gather data on all indicators contained in the eLearning data plan took much longer than anticipated. The teacher and learner level WAN usage data for example required additional funding to be made available which led to this data only being available for the last 15 months of the Game Changer period.
- The DSU ultimately covered the cost to collect the teacher and learner questionnaire data. No budget was made available for data collection. However, the WAN data came with a cost attached to it and the WCED secured the budget for this data.

- The Game Changer methodology is reliant on the availability of timeous data. Ensuring frequent access to data was challenging. Initially, only the output and implementation data were available for reflection and decision-making. Data on outcomes only followed later. Once the outcome data became available the conversations at stocktake meetings started shifting. At first, officials questioned the amount of data being collected and battled to understand how it all fit together. Once the outcome data became available a more coherent picture emerged of the performance of the Game Changer which motivated officials and sample schools to continue with the data activities.
- Not all eLearning Game Changer indicators were ultimately measured. The reasons for this ranged from system limitations (ePortal data), limited capacity of subject advisors (classroom observations) and eLearning advisors (timetable schedule) to collect the data. This led to a greater reliance on the survey data to fill some of the data gaps.
- The validity of some of the indicators is questionable. Initially ICT integration was to be measured by a teacher competency assessment tool. This tool was developed by WCED but focused more on teachers' ability to integrate ICT as opposed to the extent they were integrating ICT (which is a behaviour change). It was therefore decided to use the teacher survey data to measure ICT Integration. A cursory overview of the literature on this topic indicates the complexity of measuring "ICT integration" and more work is needed to ensure this construct is properly measured.

Given the identified shortcomings, it was deemed appropriate to conduct a high-level assessment of the quality of the eLearning Game Changer data according to the following criteria: completeness, uniqueness, validity, accuracy, timeliness and consistency (Table 26). The data quality assessment produced a "confidence barometer" of the data being presented. The four tiered assessment framework worked well in flagging the areas of concern and allowed the stakeholders to proceed with caution when interpreting some of the data - especially data categories or tools with a red or amber red rating[14] attached to it (i.e. the WAN Usage data, time table data, ePortal data and teacher training data as per Table 26)

[14] Amber red denoted by orange colour; amber green denoted by light green colour

Table 26: Game Changer data quality assessment

Data category/ Data tool	Complete	Unique	Timely	Valid	Accuracy	Consistent	Overall
Survey data	1	2	3	4	5	6	
WAN usage	7	8	9	10	11	12	
ePortal usage statistics	13	14	15	16	17	18	
Timetable schedule	19	20	21	22	23	24	
Learner reporting sheet	25	26	27	28	29	30	
Delivery checklists	31	32	33	34	35	36	
Teacher training report	37	38	39	40	41	42	
CeI WAN downtime report	43	44	45	46	47	48	
CeI Service desk report	49	50	51	52	53	54	

(DSU, 2019a)

A short description of the six dimensions is provided below

- Completeness: refers to the proportion of data collected in relation to the target population/ target group.
- Uniqueness: ensuring that no record is duplicated. The use of unique identifiers in the Game Changer dataset largely mitigates the challenges (CEMIS number, PERSAL number and education management information system (EMIS) numbers are used throughout).
- Validity: is the extent to which the indicator or data is a true reflection of the construct being measured. Many different types of validity exist and typically requires in-depth statistical analysis to calculate.

- Accuracy: is very closely linked to validity. It refers to the extent that the data represents or are a true reflection of the real world. Human error and system challenges can compromise accuracy of data.
- Timeliness: the extent to which the data is representative of the reality at a given point in time. Quick turnaround of data is the key consideration here.
- Consistency: is the absence of huge differences or fluctuations when comparing the same thing over time.

It must be kept in mind that the data quality assessment was undertaken against the understanding that for some data sets (for instance WAN usage data and questionnaire data), data collection only occurred for a select set of schools. "Completeness" as a dimension is therefore assessed against the selected set of schools, not the full set of public ordinary schools in the province.

Instead of performing the assessment on each data collection tool, some data collection tools were combined into data categories as similar principles would apply. The notes to table 26 specify the instances where the data collection tools were combined into a data category. In some cases, it was necessary to do the data quality assessment at the data collection tool level, given the peculiarities of the tool (i.e. CeI service desk data, WAN Usage data, ePortal usage statistics, timetable schedule, learner reporting sheets, teacher training report and WAN downtime report).

The numbering inside the table is linked to the notes below, and explains the rating assigned. The table does not reflect on the data that did not materialise, i.e. the classroom observation data, Moodle statistics and the eAdmin stream outcome data. An overall rating is provided in the far right column, which is utilised in Chapter 9.

Notes to Table 26:

Survey data (Tools included: Teacher, learner and principal questionnaires)

1. Consistently more than a 25% response rate on all surveys obtained.
2. All Game Changer data was collected using unique identifiers – learner and teacher data were collected against CEMIS and PERSAL numbers. School data was collected against EMIS numbers.
3. Survey data were analysed within 2 months of survey closing.
4. Statistical analysis of all indicators was required to determine the validity of the data – no rating therefore indicated.

5. Difficult to determine accuracy of survey data – some checks were done during the analysis process.
6. Survey data was checked for major deviations from one survey to the next. The use of unique identifiers made it easy to spot inconsistencies in the data.

Learner and teacher WAN usage report

7. WAN usage data only reflected wireless access hence not a full view of WAN usage at learner and teacher level.
8. WAN usage data was collected at learner and teacher level using the login codes. Initially the learners were using a different login number, which needed to be matched with their CEMIS number. But this was resolved later.
9. This data was available on a monthly basis and was analysed within two weeks of submission.
10. These datasets were produced automatically and hence validity issues did not apply.
11. If probes (which enables the production of the data) were switched off by mistake, the accuracy of the data were compromised.
12. Difficult to establish consistency as data will fluctuate over time. Time is needed for new systems to stabilise – hence consistency not scored.

ePortal usage statistics

13. ePortal data: the portal was expanding continuously as new resources were added. Tagging of resources was not optimally done, hence it was difficult to obtain a complete picture of the resources available. Enhancement to the system was discussed with the service provider to address tagging issues and naming conventions
14. ePortal: Due to learners not needing to login, individual usage of the ePortal system could not be established
15. Current ePortal statistics were produced on daily basis but it did not meet the eLearning data plan requirements.
16. These datasets were produced automatically and hence validity does not apply.
17. This data was produced by a system and accuracy therefore depended on the maturity of the system. New systems had been put in place in many instances, hence Amber Green rating as system challenges were being resolved.
18. With new systems time was needed for data to “settle down” and for trends to emerge. Inconsistencies were therefore anticipated during the initial phases.

Timetable data

19. Timetable data: very limited data received from sample schools.
20. All Game Changer data was collected using unique identifiers –School data (such as timetable data) was collected against EMIS numbers.
21. Timetable data collection process did not receive priority at school level resulting in irregular timeline for submission by schools.
22. Timetable: Valid to the extent that at least 70% of teachers of a particular school submitted their timetables
23. Timetable utilised an excel sheet which was not optimal – significant room for errors.
24. Timetable data: consistency questionable as much smaller number of sample schools submitted timetable data during the second round (17 schools in 2018, compared to 34 in 2017).

Learner reporting sheets

25. Learner performance: major gaps with Grade 10-12 performance data
26. All Game Changer data was collected using unique identifiers – learner data was collected against CEMIS numbers.
27. Learner performance data was received quite late - schools have a certain time period within which to submit the data, but follow-ups always needed to be done. System to support more timely submission was not yet in place.
28. Learner Performance: Valid to the extent that Maths and Language exams were measuring the actual syllabus learners were supposed to have mastered through the year. Content covered in these exams was beyond the control of the DSU.
29. This data was produced by a system and accuracy therefore depended on the maturity of the system. However, there was no standard system for grades 10-12. A new system is being developed.
30. Consistency in the fact that learners were tracked over time without numbers declining by more than 10%.

Delivery checklists (Tools included Technology roll out checklists, eCulture checklist, CEMIS system report)

31. Delivery data: the eLearning team closely monitored WAN, LAN, device roll out, signing of MOUs/ MOAs and eAdmin roll out.
32. All Game Changer data was collected using unique identifiers –School data was collected against EMIS numbers.
33. Delivery data was tracked on weekly or monthly basis. System utilised provided updates on continuous basis.
34. These datasets were produced automatically and hence validity does not apply.
35. Delivery data was easily checked and validated given that these were tangible “products” (e.g. WAN, LAN) being rolled out. Validity issues were therefore less of a concern.
36. Delivery data: it was easy to pick up inconsistency as one would focus on the deviation between targets and actual delivery.

Teacher training report

37. Data collection included all participants i.e. PERSAL and School Governing Body teachers. There was however no system currently which integrated all training data.
38. All Game Changer data was collected using unique identifiers – learner and teacher data was collected against CEMIS and PERSAL numbers (SGB - Temporary PERSAL or Identity number).
39. Long waiting time for teacher training data – system was under development. A skills development data capturing system is under development by the WCED human resource department. Phase 1 is complete.
40. Teacher Training Data: Valid to the extent that teachers who attended /did not attend training courses have been accurately captured. The lack of systems for teacher data raised questions about the validity of this data.
41. Teacher training system was not in place, making it extremely challenging to determine the accuracy of the information (validity of PERSAL numbers and ensuring the data could be

validated against teacher data that existed elsewhere in the system was impossible to determine).

42. Due to the absence of a system, we were unable to determine the reliability of the data.

WAN downtime report

43. WAN downtime was regularly measured through probes installed in 111 sample schools.
44. All Game Changer data was collected using unique identifiers –WAN downtime reflected via school EMIS numbers.
45. WAN usage data produced on monthly basis, with analysis performed within two weeks of submission.
46. These datasets were produced automatically and hence validity was not a major concern.
47. If probes (which enables the production of the data) were switched off by mistake, the accuracy of the data were compromised.
48. With new systems, time was needed for data to stabilise. Inconsistencies were therefore anticipated during the initial phases.

CeI service desk data

49. Service desk data was continuously improved on, to reflect all queries logged by schools. Given the multiple players involved (schools, district, and service providers) some challenges were encountered to obtain a complete view of the total queries.
50. Collected at school level, using EMIS number.
51. This data was produced monthly and analysed within 2 weeks of submission.
52. The dataset was submitted automatically so data is valid to the extent that the system produces an accurate view of helpdesk queries.
53. The CeI system was significantly expanded during the Game Changer period. The amber green rating reflects the early challenges, for e.g. categorising queries in a meaningful way and ensuring an accurate reflection on the status of a query (resolved, cancelled, open etc.).
54. Difficult to establish consistency as data will fluctuate over time. Time is needed for new systems to stabilise – hence consistency was not scored.

8.7 Summary of modification to Deliverology framework and the gains of these modifications

The standard Deliverology approach does not have a dedicated step assigned to measurement – instead it is addressed under various Deliverology steps: the selection of goal level “metrics” is dealt with as part of Step 1, while the issue of target setting forms part of Step 3. Metrics, the terminology used by Deliverology is defined as part of step 2 but fall short on providing a proper definition. Even though Deliverology suggests the use of different indicator categories, limited guidance is provided on how the indicator selection should be done. Implied in the Deliverology literature is that existing indicators will be used to track performance. The above is not surprising given that according to Deliverology, the data collection should be undertaken by the implementing departments.

The Western Cape DSU deviated from this due to limited data capabilities, and resources within the implementing departments to perform the required performance measurement activities. Performance measurement is a highly technical endeavour and includes indicator selection and construction, deciding on the appropriate data collection method, designing new data collection tools where required and putting systems in place to collect the data. The DSU had oversight of all performance measurement aspects, with a specific focus on indicator construction and formulation due to its technical nature.

Drawing on the Western Cape experience, an additional step was introduced to include the development and application of indicators. This chapter provided a theoretical background to indicator development, clarifying the meaning of an indicator, as well as tracing the historical roots of different indicator categories. The categorisation of indicators provided a framework against which to discuss the eLearning Game Changer indicators. Three categories of indicators were distinguished: indicators that reflect the level of measurement, indicators that reflect the complexity of the concepts being measured and indicators that capture the level of the intervention (in this instance, a social programme). A mix of qualitative, quantitative, single, aggregated, and composite indicators were required to measure the eLearning theory of change (i.e. the outcomes) and programme implementation. Several examples were provided of how the challenging concepts (or constructs) were conceptualised and operationalised drawing on the three indicator categories.

The gains of adding the formulation of indicators as a dedicated step are threefold. Firstly, it ensures a more deliberate consideration of the type of indicators best suited for measuring the theory of change as opposed to opting for indicators based on data availability and what is easiest to measure. Options to consider include quantitative versus qualitative indicators, single versus composite indicators and output versus outcome indicators. Secondly, the expanded step helps to addresses the common challenges associated with social measurement, in particular the measurement of outcomes which are often intangible and more difficult to measure. Through dedicated processes of conceptualisation and operationalisation these intangible phenomena are properly operationalized into quantifiable and more measurable elements. Finally, the benefit of following a systematic process to indicator development ensures appropriate data collection methods and tools can be identified.

Contrary to Deliverology's recommendation that all goals (outcomes) should have targets, the DSU did not set outcome targets. Technically, targets could have been set for all the outcomes following the first measurement, but it was decided to utilise the Game Changer period to establish an evidence base around eLearning, so that meaningful target setting could be done in future.

The indicator development and execution are undertaken as part of the planning phase and is therefore well suited to form part of Step 3 of Deliverology (Plan for Delivery). To capture the additional sub-step (and its accompanying elements) as well as the changes suggested in previous chapter, I propose that Step 3 be changed as follow – with bold denoting the additions:

From:

Step 3: Plan for delivery:

- Sub step 3a: Determine your reform strategy which include priority strategies milestones, people responsible and the resource requirements.
- *Sub step 3b: Draw the delivery chain (not covered in this study)*
- Sub step 3c: Set targets and establish trajectories

To:

Step 3: Plan for delivery

- Sub step 3a: Determine your reform strategy **(addition of logic model for performance monitoring)**
- **Sub step 3b: Develop the indicators**
- *Sub step 3c: Draw the delivery chain (not covered in this study)*
- Sub step 3d: Set targets and establish trajectories **(where appropriate)**

Chapter 9: Drive Delivery

9.1. Introduction

Step 4a of the Deliverology framework - establishing routines to drive and monitor performance – consists of two elements: (a) putting in place routines to drive delivery and (b) measuring the performance of the selected priority programmes. In this chapter we discuss Step 4a, and the value add of drawing a clear distinction between performance monitoring and outcome monitoring in reporting the findings of the eLearning Game Changer.

We do not necessarily presume a direct causal link between the modifications that were made to the Deliverology approach and the positive results of the eLearning Game Changer as outlined in this chapter. If this had been the purpose, a full assessment of the DSU 4P model (Priority, People, Performance and Problem solving) would need to be undertaken. Having said this, we would argue that the modifications to the performance measurement related elements of Deliverology did in fact result in strengthening the monitoring process and introduced new demands for accountability and performance that contributed (at the very least) to the eventual successful outcomes that were observed in the eLearning Game Changer.

I begin by describing the context within which the performance monitoring took place and specifically focus on the routines that were established to accommodate discussions around the performance of the Game Changer. Following on from this I then discuss the monitoring of performance as per the standard Deliverology framework before elaborating on the modifications. I will draw on the eLearning output and outcome data to demonstrate the practical results of these modifications.

9.2. Establishing routines to drive performance

Routines are a critical element of the Deliverology approach, as they provide opportunities for the leadership to engage with the progress against the selected priorities and to address problems as they arise. To ensure the routines are focused on the status of delivery, a variety of tools were developed as

part of the Deliverology approach. These include stocktake reports[15], trajectories[16], project memos[17], "deep dives"[18] and league tables[19] to summarise progress against a variety of elements.:

All the reporting tools utilise a four-tiered colour assessment framework. The four-tiered assessment framework was applied as follows:

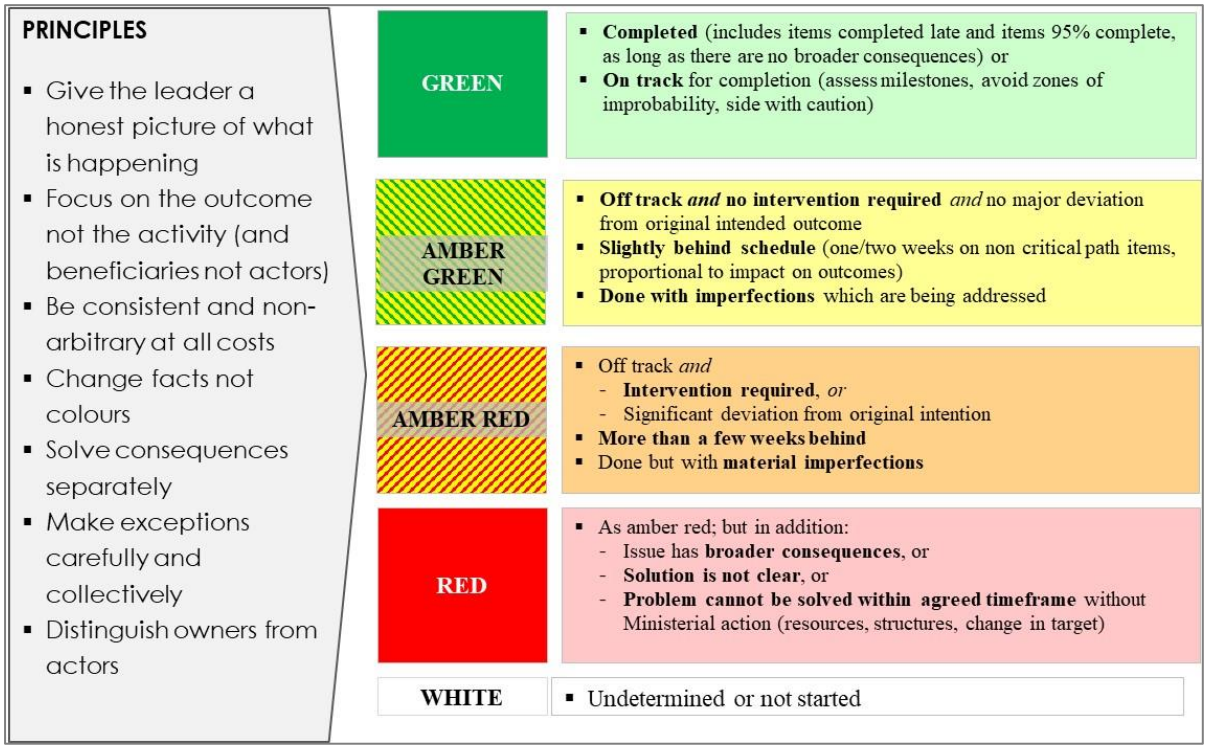


Figure 26: Four-tiered assessment framework

(Delivery Associates, 2015)

[15] Stocktakes are regular meetings with political and administrative heads where progress reports are presented, and blockages resolved.

[16] Trajectories provides an estimated projection of an indicator's path over time from its baseline (starting point) to the ultimate target.

[17] Project memos are typically directed at the political head and provides a brief update on the selected priorities. It included progress against targets, areas where decisions or intervention was needed, as well as raising any other issues that would impact on delivery.

[18] Deep dives are undertaken to gain a fuller understanding of a problem. A deep dive would typically encompass fieldwork, interviews with stakeholders and the development of an action plan to address the problem.

[19] League Tables entail a ranking or assessment of delivery aspects. For example, in the "likelihood of delivery" league table issues such as the degree of challenge, quality of planning, capacity to drive progress and the stage of delivery was assessed. Footnotes 1-5: Source: Barber, Moffit and Kihn, 2011)

Delivery routines have a very specific purpose. Barber et al. (2016) are clear on what these routines are not: they are not simply staff meetings where updates on progress are provided or where data is reviewed. Instead, delivery routines must have the following characteristics and meet the following requirements:

- Delivery routines need to be regular enough to build and sustain the momentum
- The delivery routines must align and support the delivery architecture
- The right people must be involved in the delivery routines: this encompasses both the political and administrative leaders
- The focus is on performance, with significant time and effort spent preparing the required documentation. The aim is to provide a concise view of performance, that is evidence based and which assists the delivery leaders to take action and resolve blockages. Every report must have a clear objective and storyline
- When reporting on progress, both the outcomes and implementation need to be covered

The WCG DSU utilised stocktake meetings as their main delivery routine mechanism. To maintain momentum, stocktakes were convened every 2-3 months. This allowed enough time between meetings to get on with delivery, while also providing frequent check in points to discuss any blockages or challenges to delivery. The table below shows the frequency of the eLearning Game Changer stocktake meetings between 2016 and 2019. Five stocktakes were convened in 2016, six stocktakes in 2017, four stocktakes in 2018 and two in 2019 (due to the end of the political term in May 2019). Over the four years, the Premier did not miss a single eLearning stocktake meeting.

Table 27: eLearning stocktake dates: 2016-2019

Year/ Month	**Jan**	**Feb**	**March**	**April**	**May**	**June**	**July**	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**
2016		2	15		31			30			29	
2017		7		19		6		15		17		7
2018		20		10			24			9		
2019		28		16								

Preparing for stocktake meetings typically involve five steps (Figure 27).

Figure 27: Steps followed by DSU in preparing a stocktake report

The first step in preparing a stocktake report is gathering the evidence with the aim of formulating a view on the progress of the priority programme (Game Changers). This assists in formulating specific areas of focus for the upcoming stocktake meeting, ensuring the right things are prioritised in the stocktake report. The stocktake report is then developed and presented to the political and administrative leadership in advance of the stocktake meeting.

As described in Chapter 4, the DSU instituted a three-tiered governance structure (Figure 28). MANCO meetings (denoted by “2”) were scheduled to coincide with stocktake meetings (denoted by “3”), for the stocktake report to be vetted beforehand. The resolution of blockages in advance of the stocktake meetings was an important function of the MANCO. The MANCO was a constituted structure and was chaired by the head of the WCED in the case of the eLearning Game Changer. The MANCO included representation from senior management officials from WCED and CeI as well as representatives from the DSU. The operational level of the governance structure is found at the bottom of the graph (denoted by “1”): the eLearning operational team in WCED and CeI were responsible for ensuring the day to day implementation of the six work stream activities. The eLearning operational team worked closely with the DSU team, hence its parallel positioning in Figure 28.

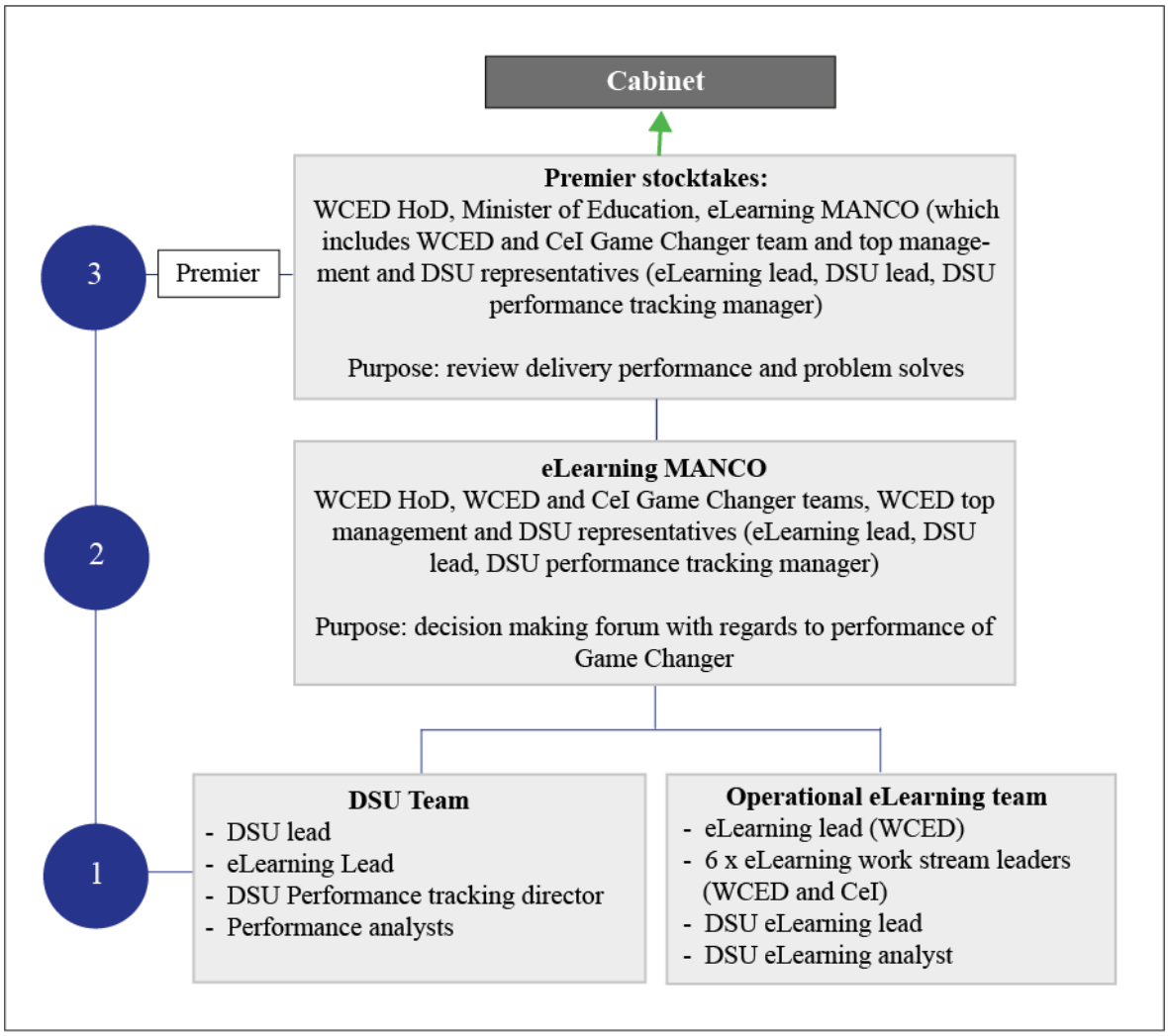


Figure 28: eLearning governance structure

(Adapted for eLearning Game Changer, Cargill, 2018)

The eLearning stocktake report was always a collaborative effort between the DSU and WCED. The process commenced approximately four weeks before a stocktake meeting with a series of internal DSU meetings to plan the content of the report based on the available data and key delivery issues at that stage. This was then followed by the MANCO meeting (more or less two weeks before a stocktake) followed by the presentation of a third draft of the stocktake report to the Education minister a week before the stocktake. Final changes would follow from the Minister meeting, ensuring the final stocktake report is circulated to all attendees 48 hours in advance of the stocktake meeting. These pre-stocktake meetings were a critical component of the governance arrangements. The management executive and the Minister wanted to be fully appraised of the issues before stocktakes and did not want surprises. In addition, these meetings gave them the opportunity to resolve problems themselves and as such added pressure on delivery.

As a final step, the stocktake report is presented to the political and administrative leaders during the stocktake meeting. Even though stocktake meetings are viewed as a problem solving forum, solutions to the problems need to be drafted and identified beforehand; with the stocktake meeting then used to make decisions on the way forward.

In the next section I discuss the performance monitoring of the eLearning Game Changer starting with the manner in which monitoring is done as per the standard Deliverology framework.

9.3 Monitoring as per the "standard" Deliverology framework

Monitoring performance is the crux of the Deliverology approach, with performance data tracked on a continuous basis to check progress against targets and to problem solve. Deliverology does distinguish between tracking the implementation of the plan and intervention outcomes but – as indicated before - does not make an explicit distinction between outputs and outcomes, neither between short- and medium-term outcomes. Within the performance measurement tradition performance monitoring is described as extending beyond outputs to also include the tracking of short-term and medium-term outcomes of program activities (Wholey & Hatry, 1992). It is also common in mainstream programme evaluation literature (Rossi et al., 2004) to make a distinction between process and outcome monitoring, with the view of improving the implementation or changing course as needed. This is often linked to a formative evaluation purpose.

Within the eLearning Game Changer performance monitoring was undertaken over the three-year Game Changer period, mainly with a formative purpose in mind. Milestones and outputs were tracked continuously to timeously identify any deviations from the set timelines and targets. The short-and medium-term outcomes were also monitored continuously to ascertain whether positive changes were occurring. Thus, the value in making this distinction proved useful in two ways: by delineating the outputs from the outcomes we prevented an over-emphasis on the easily measurable aspects (outputs), as well as ensured realistic expectations were set in terms of the achievement of outcomes. Given the short time frame, the emphasis needed to be on short term outcomes, more so than the medium term outcomes.

I discuss both the performance monitoring and outcome monitoring in the eLearning Game Changer in detail below.

9.4 Performance monitoring of outputs and milestones in the eLearning Game Changer

9.4.1 Introduction to performance monitoring of outputs and milestones

Monitoring the performance of outputs and milestones in the eLearning Game Changer was undertaken in the following way:

i) Reporting on outputs against targets
ii) Reporting on the milestones against set deadlines

In both instances, the four-tiered assessment framework was utilised (Ref. figure 26). The link between the outputs and accompanying milestones is demonstrated by way of an example from the eTechnology work stream. An annual output target was set in terms of learner devices in year 1 (2016), with milestones reflecting the specific actions that had to be undertaken in support of ensuring the annual target was achieved. Every milestone had a due date and is indicated in brackets in the table below.

Table 28: Link between outputs and milestones for the Learner device roll out in the eTechnology work stream

Learner devices outputs: 2016/17	Accompanying milestones
8000 learner devices rolled out to model schools	Identify budget to fund learner devices for model schools (Due date: 8 Feb 2016)
	Identify possible sources of sponsorship / donations / other projects for shared devices in enhanced schools (Due date: 10 February 2016)
	Develop implementation plan to source devices aligned to eLearning Game Changer requirements for enhanced schools, including design and development of marketing material (Due date: 22 Feb 2016)
	Finalise procurement model for learner devices (model schools) (Due date: 26 Feb 2016)
	Prepare tender documentation (Due date: 4 March 2016)
	Procurement: Advertise (Due date: 31 March 2016)
	Procurement: process and award 9Due date: 15 April 2016)

(DSU, 2016b)

Outputs with targets were typically developed at the start of the year. Figures 29 and 30 show an extract from the February 2017 stocktake where the upcoming financial year's outputs were presented – the slide covers the eInfrastructure and eTechnology work streams (slides 29 and 30 respectively).

Drawing on the figures below, one can see that 1278 schools were scheduled to receive the WAN by 31 March 2018. Colour coding of outputs was done, utilising the four-tiered assessment framework to highlight any potential challenges. In Figure 29 for example, challenges were anticipated with the WAN roll out for 191 schools due to the remoteness of these schools. The red colour coding means other solutions had to be found for these 191 schools. The green blocks in these figures indicate that the team did not foresee any challenges with these outputs and were confident that the targets would be achieved. The white block for the 'Slim Lab refresh" project (Figure 30) means this project had not commenced at that stage and therefore no rating could be assigned.

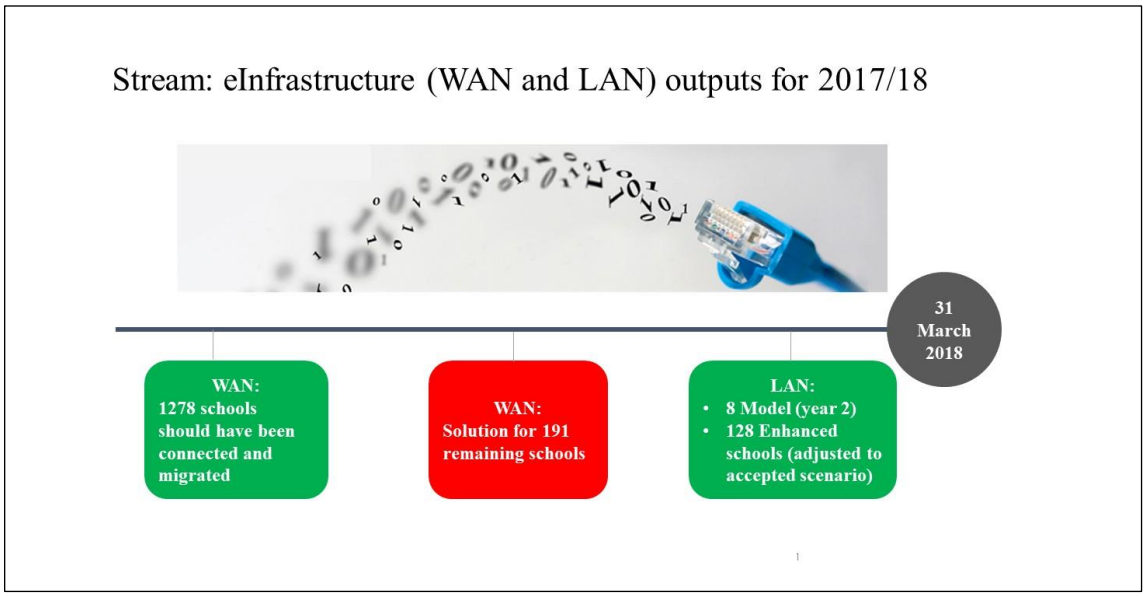


Figure 29: eInfrastructure outputs for 2017/18

(DSU, 2017a)

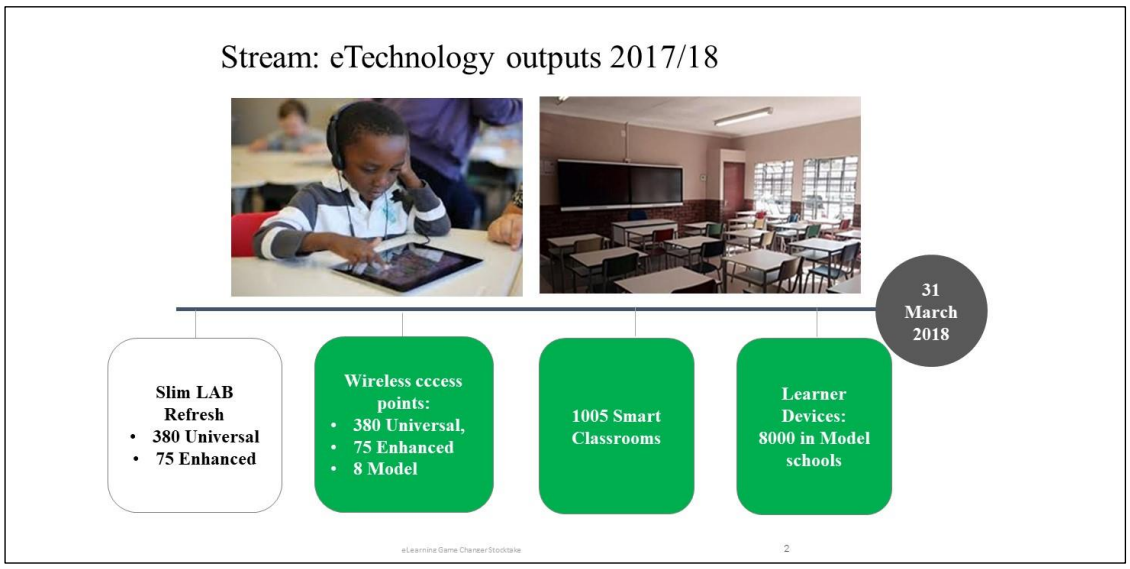
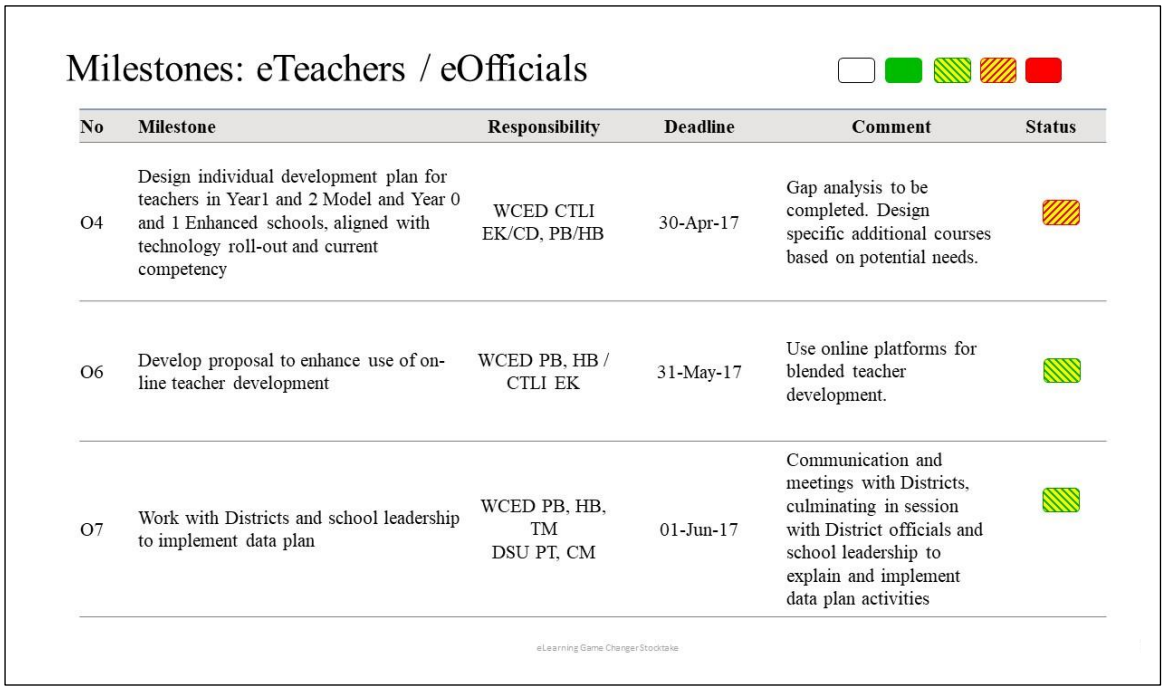


Figure 30: eTechnology outputs for 2017/18 (DSU, 2017a)

Reporting on milestones was done by providing a status update on milestones per work stream. Unlike outputs, milestones were not always developed for the year ahead, but expanded and changed as roll-out progressed. Figure 31 provides an extract of the milestone feedback for the eTeachers/eOfficials work stream. A short description of the milestone was provided, followed by the organisation and person responsible (initials indicated). In some instances, a milestone was the responsibility of multiple officials, which was the case for all three milestones shown in figure 31. Every milestone had a deadline attached to it – and for the purpose of this example, the milestone update was provided on the 6th of June 2017. The milestone colour rating was done on the basis of the deadline – if significantly behind schedule or a blockage was being experienced, it would receive an amber red rating (e.g. milestone no 4 as per Figure 31 was due for completion at end of April but which meant it was two months behind schedule). The comment block explains why a specific colour rating was provided and should indicate why the deadline had not been achieved. The two amber green ratings for milestones 6 and 7 indicate that these milestones were slightly delayed, but that no intervention was needed in completing the milestones. In all instances, evidence must be available to motivate the colour of a milestone, i.e. for completed milestones evidence must be submitted that demonstrate the completion of the milestones.



Milestones: eTeachers / eOfficials

No	Milestone	Responsibility	Deadline	Comment	Status
O4	Design individual development plan for teachers in Year1 and 2 Model and Year 0 and 1 Enhanced schools, aligned with technology roll-out and current competency	WCED CTLI EK/CD, PB/HB	30-Apr-17	Gap analysis to be completed. Design specific additional courses based on potential needs.	
O6	Develop proposal to enhance use of on-line teacher development	WCED PB, HB / CTLI EK	31-May-17	Use online platforms for blended teacher development.	
O7	Work with Districts and school leadership to implement data plan	WCED PB, HB, TM DSU PT, CM	01-Jun-17	Communication and meetings with Districts, culminating in session with District officials and school leadership to explain and implement data plan activities	

eLearning Game Changer Stocktake

Figure 31: Examples of milestone reporting: eTeachers/eOfficials (DSU, 2017b)

9.4.2 Value add of measuring outputs separately in the eLearning Game Changer

The primary value addition in distinguishing output monitoring from outcomes monitoring is that it assists in identifying the reason for substandard or non-performance of social interventions. With output monitoring the aim is to establish whether implementation occurred as designed and planned (the 'fidelity rule'), and if not, what must be done to rectify this. In the absence of continuously checking whether implementation has occurred as planned, an incorrect conclusion can be reached that the intervention is unsuccessful in shifting the outcomes. This is called theory failure, i.e. the stipulated causal pathways as captured in the theory of change is faulty and do not produce the desired results. With implementation failure, the delivery deviated from the plan. This distinction is important as different actions are required to rectify implementation failure compared to theory failure.

In demonstrating this point, Table 29 presents a summary of the performance indicators per work stream, indicating the data quality (column 2), the actual performance vs three year target (column 3) and an achievement rating based on column 3's input (column 4). The data quality rating explained in chapter 8, is carried over into this table. It enables the reader to assess the achievement of outputs, while keeping the quality of the data in mind. For example, indicators with an amber red or red data quality rating should be interpreted with caution as it means that this data was of poor quality. The unit of analysis is indicated in brackets in the target column.

The achievement rating (column 4) was done based on the following criteria:

- Green = target satisfactorily met (90%+ of outputs achieved))
- Orange = target met to a moderate degree of achievement (70% to 90% of outputs achieved)
- Red = target not met (less than 70% of outputs achieved)

It is important to state that the 'targets' included in the table below were in fact 'moving targets' and reflect the actual situation in 2019. Targets were adjusted from year to year, depending on budget and improved information on what was taking place on the ground.

Table 29: Achievement of outputs per work stream

Performance Indicators	Data quality rating	Target (where applicable)	Achievement rating
eInfrastructure			
Number of schools connected to broadband		1250/1278 (schools)	97%
Number of enhanced schools connected with LANs		435/366 /schools)	118%
Number of universal schools with wireless access points		873/900 (schools)	97%
Number of enhanced schools with wireless access points		72/72 (schools)	100%
Number of queries logged at CEI service desk		No target	Not applicable
eTechnology			
Number of enhanced schools with slim laboratory refresh		1160/1160 (schools)	100%
Number of universal Schools with slim laboratory refresh			100%
Number of smart classrooms (teacher devices) deployed at enhanced schools		7530/7530 (classrooms)	100%
Number of smart classrooms (teacher devices) deployed at model schools			100%

Performance Indicators	Data quality rating	Target (where applicable)	Achievement rating
Number of learner devices that has been distributed to model schools		16183/16183 (devices)	100%
eTeachers			
Number of principals and teachers trained in basic ICT competencies		Unknown/32214	Not applicable
Number of teachers trained in ICT Integration		Unknown/ 4056	Not applicable
Number of school management officials trained in ICT integration		Unknown/1239	Not applicable
eContent			
Number of topics with digital resources available		Poor data quality	Not applicable
eAdmin			
Number of model schools with web-based school admin systems in place		16/16 (schools)	100%
Number of enhanced schools with web-based school admin systems in place		45/50 (schools)	90%
Number of universal schools with web-based school admin systems in place		22/25 (schools)	88%
Number of model schools where data dashboards have been made available/ rolled out		16/16 (schools)	100%
Number of targeted enhanced schools where data dashboards have been made available/ rolled out		150/150 (schools)	100%
Number of model schools with learning management system implemented		4/16 (schools)	25%
Number of enhanced schools with learning management system implemented		0/10 (schools)	0%
Number of model schools with parent/learner portal implemented		0/16 (schools)	0%
Number of enhanced schools with parent/learner portal implemented		0/10 (schools)	0%

Performance Indicators	Data quality rating	Target (where applicable)	Achievement rating
eCulture			
Number of schools that sign MOU or MOA or Letter of commitment		1043/1499 (schools)	70%

Table 29 provides a predominantly positive assessment of the achievement of outputs. However, some issues regarding data quality prevented us from making a comprehensive assessment of all the performance indicators. These limitations are elaborated on below:

- Most of the targets for the eInfrastructure and eTechnology work streams were achieved, but with roll out continuing until end of March 2019, not all targets were reached at the time of compiling the February 2019 stocktake report. No target was set for the number of queries logged with the CeI service desk and hence no value is reported in the table. This data was however analysed continuously and broken down into ICT categories to determine the most prevalent problems.
- The work under the Digital content stream (eContent) was geared towards growing the digital resource base for Maths and Language primarily but not solely for these subjects. The ePortal, an online platform launched by WCED in September 2015, was viewed as the mechanism by which learners and teachers would gain access to these digital resources. Unfortunately, much of the ePortal content had not been adequately quality-checked or tagged, leaving it to the user to decide on the most relevant and useful content. The digital content challenges were being addressed, but for the purpose of reporting on the eContent work stream during the Game Changer period, data quality was extremely poor making it impossible to track the availability of digital content in a meaningful way.
- Tracking the number of teachers and school management officials being trained (eOfficial stream) was challenging given the lack of an integrated training database where all training providers' data was stored centrally. This lack of an integrated database was identified as a major shortcoming, as this prevented the WCED from keeping track of each individual staff member's training (whether it be teachers/ officials/ management staff). Given this, the data quality was rated as amber red. Some data was presented at the final stocktake in February 2019 but due to the aforementioned limitations a comprehensive view of ICT and ICT Integration training undertaken by teachers and management staff could not be provided.
- The majority of the eAdmin stream's deliverables were met except for the learning management system outputs and the learner/parent portal outputs. A change in the roll-out of the eLearning

management system towards the end of 2018, necessitated a re-thinking of the targets, and hence these targets were not achieved by the time the Game Changer period was concluded. The roll-out of a learner/parent portal was always intended to take place after the Game Changer period, but the eAdmin team wanted these outputs noted in the roadmap.

- The eCulture output entails all principals signing a MOU/ MOA or a letter of acknowledgement, detailing their role in the roll-out of the eLearning Game Changer. As at the end of January 2019, 1043 out of 1499 schools had signed the letter.

In conclusion: I would argue that the single biggest gain of a dedicated focus on performance monitoring is that it allows one to readily identify any deviation from the planned delivery of the intervention. Although this is a well-known 'truism' in standard programme evaluation literature, this is not the case in the standard Deliverology framework. Drawing on the eLearning Game Changer data as presented in Table 29 it is evident that in the main implementation happened as planned. The under-performance in the eAdmin work stream does not constitute implementation failure, as the reason for the non-achievement of performance targets related to a deliberate decision to delay implementation. However, a continued delay in implementing the planned projects will ultimately impede on the achievement of the eAdmin work stream outcomes and must continue to receive attention. The lack of good quality data for two of the eLearning Game Changer work streams (digital content and eOfficials) is a cause for concern. In the absence of data, delivery cannot be measured.

Overall, the output data presented for the eLearning Game Changer demonstrate the value of ensuring a dedicated focus on implementation. We now move to the outcome monitoring of the eLearning Game Changer and the benefit of distinguishing this from performance (output) monitoring.

9.5 Outcome monitoring in the eLearning Game Changer

This discission is devoted to two issues:

- An overview of the outcomes data collected and how it was reported utilising an ACCESS, ACCEPTANCE and USE framework
- Value add of measuring the eLearning Game Changer outcomes separately from the outputs

9.5.1 An overview of the outcomes data collected and reporting thereof

I list the data collection tools that were used to collect data on the outcome indicators in Table 30. I also include the period and/ or frequency of data collected, the number of schools covered by the data collection and the level at which the data was analysed.

The survey data (for learners, teachers, and principals) were collected at specific times in the year, while the WAN and CeI service desk data were collected monthly between March 2018 to February 2019. The timetable data and classroom observation data, which did not materialise, were scheduled for periodic collection: annually in the case of the timetable data and quarterly in the case of the time classroom observations. As explained in Chapter 5, different groups of schools were covered through the various data collection efforts. The teacher and learner survey activities took place in the 63 eLearning sample schools, while the WAN downtime and WAN usage data were collected for all 111 sample schools. The principal questionnaire and CeI service desk data were collected for all public ordinary schools. Learner recording and reporting sheets containing attendance and performance data were collected at year end (2017 and 2018).

Table 30: Planned outcome data collection

Data collection methods and tools	Sampling/selection of schools (cases)	Time period of data collection	Level of analysis
Surveys			
Learner questionnaires	63 schools	• July 2017 (Model schools only) • Nov 2017 • April 2018 • Nov 2018	Learner, using CEMIS number
Teacher questionnaires	63 schools	• Nov 2017 • April 2018 • Nov 2018	Teachers, using PERSAL number
Principal questionnaire	All schools	• 2017 (March) • 2018 (March)	School, using EMIS number
Timetable schedule	47 schools	2017 and 2018	Teachers, using PERSAL number
Observations			
Classroom observation schedule	16 schools	Quarterly in 2018	Teachers, using PERSAL number

Data collection methods and tools	Sampling/selection of schools (cases)	Time period of data collection	Level of analysis
Platform data			
Learner and teacher level WAN usage report	111 schools	March 2018-Feb 2019 (Monthly)	Individual (converting Broadband login to CEMIS and PERSAL number)
WAN Downtime report	111 schools	March 2018-Feb 2019 (Monthly)	School, using EMIS number
CEMIS report	All schools	September 2018-Feb 2019 (Quarterly)	School, using EMIS number
CEI Service desk report	March 2018 – Feb 2019	March 2018 – Feb 2019 (Monthly)	School, using EMIS number
Learner data			
Learner recording and reporting sheets	111 schools	2017 and 2018 (Annually)	Learner, using CEMIS number

The teacher and learner data collection processes were explained in Chapter 5 but are briefly revisited here as it constituted the most significant data collection activities.

On the learner side, two learner groups completed the questionnaires. In 2017, when the data collection started, grades 4,6,8 and 10 were included. This is referred to as the "2017 cohort". This same group of learners completed the questionnaire again in 2018 when they progressed to the next grade. In 2018, another cohort (2018 cohort) came on stream and completed the learner questionnaire in April 2018 and November 2018. The bottom of table 32 provides a breakdown of the targeted grades for the two learner groups as well as the frequency of data collection:

- The 2017 model school learners in targeted grades completed the questionnaire three times: August 2017, November 2017 and November 2018 (when they progressed a grade)
- The 2017 enhanced and universal school learners in targeted grades completed the questionnaire twice: November 2017 and November 2019 (when they progressed a grade)
- The 2018 group of learners (model, enhanced and universal) completed the questionnaire twice: April 2018 and November 2018

The completion rates for the learners, broken down at district level, are also provided in Table 31. Both the number of learners completing the questionnaires, as well as the proportion of learners completing the questionnaires per district are shown in table 31. Overall, high completion rates were achieved for all periods.

Table 31: Completion rates: Number of learners completing questionnaire over time

District response rate	Aug 2017	Nov 2017	April 2018	Nov 2018
Cape Winelands	526 (73%)	2261 (69%)	2841 (79%)	4222 (63%)
Eden and Central Karoo	611 (76%)	2505 (76%)	1816 (52%)	6055 (92%)
Metro Central	724 (79%)	2542 (93%)	2037 (73%)	3790 (73%)
Metro East	404 (79%)	1564 (77%)	1996 (87%)	3341 (80%)
Metro North	858 (77%)	1961 (58%)	1725 (49%)	4829(73%)
Metro South	644 (79%)	1185 (43%)	926 (31%)	1879(33%)
Overberg	711 (79%)	1152 (66%)	1024 (56%)	2134 (64%)
West Coast	959 (76%)	2054 (93%)	1843 (84%)	3452 (86%)
TOTAL	**5437 (77%)**	**15224 (71%)**	**14208 (63%)**	**29702 (70%)**
Cohort group grade breakdown				
Cohort 2017 (Learner Group)	Grade 4,6,8,10	Grade 4,6,8,10		Grade 5,7,9,11
Cohort 2018 (Learner Group)			Grade 4,6,8,10	Grade 4,6,8,10

Teachers in the 63 eLearning sample schools completed the questionnaire three times: November 2017, April 2018 and November 2018. Table 32 provides a summary of the teacher completion rates for the three time periods.

The teacher completion rates stayed relatively stable, however the number of teachers completing the questionnaire increased quite substantially between 2017 and 2018 (Table 32). The reason for this is that WCED only has data available for the state-paid teachers, and not for teachers paid by the School Governing Body (SGB). State-paid teachers receive a PERSAL number and this number was used to produce a login code. To mitigate for the lack of information on the SGB teachers, and to ensure their participation, back up log-in codes were made available to these teachers. There was also a significant drive between November 2017 and April 2018 to encourage schools to use the back-up codes for teachers without login codes. This led to increased numbers of teachers participating in the survey activities.

Table 32: Questionnaire completion rates: teachers

Response rate	Nov 2017	April 2018	Nov 2018
TOTAL	459 (75%)	961 (64%)	1051 (70%)

In total, approximately 30 000 unique learners and 1 000 teachers completed the questionnaires over the year-and-a-half period. However, it was decided that only “matched data” would be reported at

stocktakes; hence in this study, I have followed the same route. “Matched data” entails matching learners and teachers using their CEMIS and PERSAL numbers, over the various data collection instances to ensure the same learners and teachers were tracked over time. This provided a strong basis for making judgements around the effects of the strategies and programmes as the same group of learners and teachers were analysed over time. Ultimately approximately 20 000 learners were matched across the two learner groups (2017 + 2018 learner group as shown in the last column in table 33). The same applies to the teachers: 700 teachers were ultimately matched by the time the questionnaire was administered for the last time in November 2018.

Table 33: Summary of “matched data” over time

	Aug 2017	**Nov 2017**	**April 2018**	**Nov 2018**
Cohort 2017 (Learner group 1)	4157	14632		10020
Cohort 2018 (Learner group 2)			14208	10708
Teachers		651	658	700

Outcome data collection started mid-2017, with the first coherent feedback on outcome data taking place at the February 2018 stocktake. In deciding how best to present the data, I developed an outcomes-based framework around ICT access, ICT acceptance and ICT use. This framework is basically a reconfiguration of the theory of change, as can be seen from the underlined text in the text box below (figure 32), but also expands on the theory of change in that the attitudinal aspects were included to fully reflect the work of the eCulture work stream.

eLearning Theory of Change
IF
Principals support the eLearning Game Changer
AND
Schools and teachers have better connectivity
AND
Learners and teachers have better access to technology (smart classrooms, learning devices)
AND
Schools receive better technological support
AND
Teachers and principals are aware of, and receive the ICT support available from the district
AND
Curriculum support officials (including planners) provide better ICT support to teachers and principals
AND
Teachers use available technology in their teaching practice, and integrate available technology and digital resources into their teaching environment
AND
Learners and teachers have better access to digital resources
AND
Targeted principals have access and use the digital school admin systems and school dashboards
THEN
Learners will be more motivated to attend school, learners will be engaged during classroom teaching and learner performance will improve

Figure 32: eLearning Theory of Change (underlined text indicating link to outcomes-based framework)

The outcomes based framework is shown in Figure 33. The three intertwined blocks demonstrate the (hypothesised) interdependency that exists between access, use and acceptance of ICT. The benefit of the reconfigured version of the theory of change was two-fold: it provided a clear emphasis on the short term versus medium term vs impact statements. All outcomes related to "access" and the additional element of "acceptance" constituted the short term outcomes, while the outcomes around ICT integration and use were defined as the medium-term outcomes. The underlying logic is that only once access is in place, will ICT be used. As an abbreviated theory of change this framework postulates the following:

> IF learners, teachers and schools have ACCESS to internet connectivity, technology, digital content, eAdmin systems, AND IF learners, teachers and principals have a positive ATTITUDE to eLearning (i.e. accepts ICT) AND IF learners, teachers and schools USE the connectivity, technology and eAdmin systems THEN one will see an improvement in learner performance.

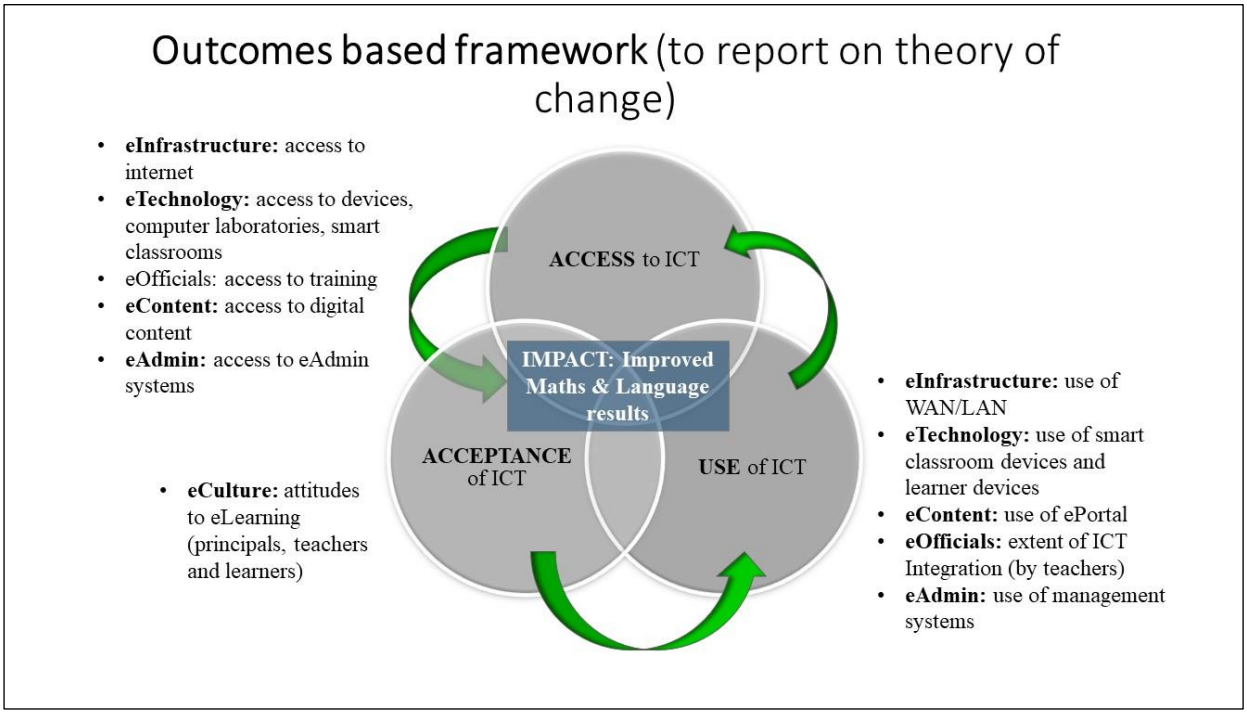


Figure 33: Outcomes based framework which aligns six work streams to theory of change

(DSU, 2018a)

Another advantage of depicting the theory of change in this manner was that the six work streams could be aligned to the outcomes based framework. With all reporting up to this point being done by way of the six work streams, this was important for continuity purposes. The eCulture stream for example is mainly addressed by the acceptance element; but instead of only measuring principal support, learners and teachers' attitudes were also tracked. The ultimate impact statement is contained at the intersecting points of the three bubbles, namely improved mathematics, and language results.

Table 34 summarises the indicators per work stream and how these align to the outcome framework of access, acceptance, and use. As discussed before, not all indicators included in the eLearning data plan ultimately materialised. The comments provided in the last column explain the additional indicators that were added to fill the gaps. Triangulation was also done by utilising the learner and teacher survey data.

The indicators that are included in the data plan, and ultimately reported on are numbered, with the numbering aligned to the data plan annexure (Annexure B1 and B2). Indicators that were introduced for triangulation purposes are indicated in grey. These indicators are not included in the data plan and

are also not reported on in this study. New indicators introduced during the Game Changer period are shown in *italics*. The amendments to the indicators reflect the dynamic nature of performance measurement, and why outcome measurement is found to be challenging. I provide further detail on some of the changes made to the original set of indicators, as well as some of the challenges encountered with the indicators.

Initially schools' data volumes as tracked through the downloads, as well as the type of data downloaded were identified to measure internet connectivity (RED in table 34). With the WAN usage data, and WAN downtime data becoming available, it was decided to replace these two indicators with more suitable indicators. The "total number of academic days affected by WAN downtime" as well as the "percentage of CeI service desk queries that relate to connectivity" were deemed to be better measures of the access to the internet than simply tracking data volumes.

New indicators were added for the eTechnology work stream to measure access to smart classrooms and computer laboratories. The reason for this is that the poor data quality (to track access to smart classrooms and computer laboratories) was evident with the first set of timetable data. There was not a proper timetable system in place, and an extremely low response rate was obtained. The decision was therefore taken to draw on the teacher questionnaire for this data. A similar reasoning applies for adding a new indicator around the teachers' access to digital content. Given the limitations of the ePortal, it was decided to rather use the teacher questionnaire data to track teachers' access to digital content.

Table 34: Indicators used to track ACCESS, ACCEPTANCE and USE

Stream	Outcome statement	Outcome Indicator	Included in analysis	Outcome framework	Reason for additional/ new indicators
eInfrastructure	Teachers have better connectivity	1.Number of teachers that connect to wireless access points to enable own device	yes	**ACCESS**	
	Teachers have better connectivity	Ranked position of "Internet Connection unreliable" as per Teacher survey	no	**ACCESS**	Triangulate findings for indicators 3 and 4
	Schools have better connectivity	Data being accessed and downloaded by schools	no		
		2. Percentage of CEI Service desk queries that relate to connectivity	yes	**ACCESS**	Replaces indicator: Data being accessed and downloaded by schools
		Schools' data volumes	no		
		3. Number of academic school days affected by WAN downtime	yes	**ACCESS**	Replaces indicator: Schools' data volumes
	Schools have better technological support	4.Number of queries reported to CEI service desk that are resolved within five working days	yes	**ACCESS INFRASTRUCTURE SUPPORT**	
eTechnology	Learners have better access to technology	5. Percentage time spent in smart classroom environment	yes	**ACCESS**	
		6. Percentage time spent in laboratory environment	yes	**ACCESS**	
		7. Number of model school learners that connect devices	yes	**ACCESS**	
		8. Number of teachers teaching in computer labs	yes	**ACCESS**	
		9. Number of teachers teaching in technology enabled classrooms	yes	**ACCESS**	

Stream	Outcome statement	Outcome Indicator	Included in analysis	Outcome framework	Reason for additional/ new indicators
eTeachers/ eOfficials	Learners have better access to technology	Number of learners that indicate they have access to computer/ tablet every day/ almost every day	no	**ACCESS**	Triangulation of indicator 5
		Ranked position of "Not enough computers for learners to use" as per teacher survey	no	**ACCESS**	Triangulation of indicator 6
	Teachers and principals are aware of the ICT integration support system in place at districts	10. Number of teachers and principals that are aware of the ICT Integration support available to them	yes	**ACCESS**	
	Teachers and principals receive ICT integration support from district	11. Number of teachers and principals that request support from district staff	yes	**ACCESS**	
	Curriculum support officials provide better ICT support to teachers and principals	12. Number of teachers and principals that are satisfied with support they receive from district staff	yes	**USE**	
	Teachers use technology in their teaching practice	13.Number of teachers that use wireless access points to enrich educational practices	yes	**USE**	
		14. Number of teachers at model schools that use learning management systems to plan and deliver online lessons	yes	**USE**	
		Number of teachers using smart classroom devices (projector, laptop, white board, visualisers)	no	**USE**	Tracked this indicator to check teachers' use of devices
	Teachers integrate technology into their teaching environment	15. Number of teachers that integrate ICT and e-resources in teaching practices	yes	**USE**	

Stream	Outcome statement	Outcome Indicator	Included in analysis	Outcome framework	Reason for additional/ new indicators
eContent	Learners have better access to digital resources	16. Number of learners that indicate they choose digital resources as a first choice when learning	yes	**ACCESS**	
		17. Count of CAPS aligned Mathematics & Language resources downloaded	yes	**USE**	
	Teachers have better access to digital resources	*18. Number of teachers indicating they have access to a variety of digital content*	Yes	**ACCESS**	Needed an indicator that would reflect teacher access due "Count of CAPS aligned Mathematics & Language resources not materialising"
	Teachers integrate digital resources into their teaching environment	19. Number of teachers that access professional learning communities on e-Portal and through other collaborative platforms on a recurring basis	yes	**ACCESS**	
		20. Number of teachers that upload content onto e-Portal	yes	**USE**	
		21. Number of teachers that share resources with other teachers within and outside of school	yes	**USE**	
eAdmin	Principals have better access to digital school admin systems and school dashboards	22. Number of model and enhanced schools that upload learner assessment results to a central repository	yes	**ACCESS**	
		23. Number of model and enhanced schools that access learner assessment and attendance data on the dashboard	yes	**ACCESS**	
	Principals use the digital school admin systems and school dashboards	24. Number of model and targeted enhanced schools that use a school admin system to collect school based assessment and attendance data	yes	**USE**	

Stream	Outcome statement	Outcome Indicator	Included in analysis	Outcome framework	Reason for additional/ new indicators
		25.Number of model and targeted enhanced school principals that report improved data collection and management (school admin system and data dashboard)	yes	**USE**	
eCulture	Principals support the eLearning GC	26.Number of school principals that express commitment to eLearning Game Changer	yes	**ACCEPTANCE**	
	Teachers are aware, understand and support the eLearning GC	Number of teachers with positive attitude to eLearning	no	**ACCEPTANCE**	Wanted to obtain data on teachers' acceptance of ICT even if not included in theory of change
	Learners are aware, understand and support the eLearning GC	Number of learners with positive attitude to eLearning	no	**ACCEPTANCE**	Wanted to obtain data on learners' acceptance of ICT even if not included in theory of change

9.5.2 The value add of measuring the outcomes separately from the outputs

There are several reasons why the assessment of outcomes should be undertaken separately from output monitoring. Firstly, without the dedicated focus on outcomes there could be a tendency to focus on the easily measurable aspects of an intervention instead of measuring what matters most (the anticipated positive benefits to the target group). Secondly, without a clear distinction between short term and medium term outcomes, programme staff may hold unrealistic expectations as to what can be achieved within a certain time period resulting in the premature discontinuation of interventions. And thirdly, the separation between output and outcome monitoring allows for the theory of change to be tested systematically and rigorously – a key tenet of theory-based evaluation approaches such as Realistic Evaluation.

I elaborate on these conclusions drawing on the eLearning Game Changer outcome data (see table 35). The outcomes are provided in the first column, followed by the list of outcome indicators as well as the baseline and end value of the indicators. The alignment with the outcomes-based framework (access, acceptance, and use) is shown in the indicator column, with the light blue cell representing the short-term outcomes.

Given that data was collected over different time periods, and at different frequencies, the baseline date and end date have been provided in brackets. No targets were set for the outcomes; hence the "change" column reflects the degree of change, based on the percentage points increase or decrease as defined in the legend below:

- Green = positive change: 5 percentage points increase or more
- Orange = no or little change: less than 5 percentage points increase or decrease
- Red = negative change: 5 percentage point decrease or more

A total of 32 indicators are included in the table below with data being available for 25 indicators.

Table 35: Achievement of outcomes as per the theory of change

Outcomes	Outcome Indicators	Data quality rating	End Value	Baseline	Change rating
Teachers have better connectivity	**eInfrastructure**				
	% of teachers that connect to wireless access point to enable own device ACCESS		53% (Q4 2018)	43% (Q2 2018)	
Schools have better connectivity	% CEI Service desk queries that relate to connectivity ACCESS		62% (April 2019)	71% (April 2018)	
	Number of academic school days affected by WAN downtime ACCESS		129 (April 2019)	131 (April 2018)	
Schools have better technological support	**ICT Support**				
	Percentage of queries reported to CEI service desk that are resolved within 5 working days ACCESS		90% (Q4 2018)	77% (Q2 2018)	
Learners have better access to technology	**eTechnology**				
	Percentage contact time spent in smart enabled classrooms ACCESS		Limited data	Limited data	
	Percentage contact time spent in laboratory environment ACCESS		Limited data	Limited data	
	% of model school learners that connect devices ACCESS		31% (Q4 2018)	28% (Q2 2018)	
	Percentage teachers teaching in technology enabled classrooms ACCESS		66% (Nov 2018)	57% (Nov 2017)	
	Percentage teachers teaching in computer laboratories ACCESS		23% (Nov 2018)	27% (Nov 2017)	
Teachers and principals are aware of the ICT integration support system	**eTeachers/eOfficials**				
	Percentage of teachers aware of ICT Integration support available to them ACCESS		79% (Nov 2018)	79% (Nov 2017)	

Outcomes	Outcome Indicators	Data quality rating	End Value	Baseline	Change rating
in place at districts					
	Percentage of principals aware of ICT Integration support available to them ACCESS		Not available	70% (March 2017)	
Teachers and principals receive ICT integration support from district	Percentage of teachers that request support from district staff ACCESS		34% (Nov 2018)	38% (Nov 2017)	
	Percentage of principals that request support from district staff ACCESS		71% (March 2018)	63% (March 2017)	
Curriculum support officials provide better ICT support to teachers and principals	Percentage of principals that are satisfied with the support they receive from district ACCESS		60% (March 2018)	50% (March 2017)	
	Percentage of teachers that are satisfied with the support received from district staff ACCESS		67% (Nov 2018)	51% (Nov 2017)	
Teachers use appropriate technology in their teaching practice	Percentage of teachers that use wireless access points to enrich educational practices USE		50% (Q4 2018)	40% (Q2 2018)	
	Percentage of teachers at model schools that use a learner management system USE		71,1% (Nov 2018)	51,4% (Nov 2017)	
Teachers integrate appropriate and available technology into their teaching environment	Percentage of Teachers that integrate ICT and eResources in teaching practices USE		30% (Nov 2018)	24% (April 2018)	
	eContent				
Learners have better access to digital resources	Percentage of learners that indicate they choose digital resources as a first choice when learning: 2017 group ACCESS		50% (Nov 2018)	42% (Nov 2017)	
	Percentage of learners that indicate they choose digital resources as a first choice when learning: 2018 group ACCESS		51% (Nov 2018)	37% (April 2018)	

Outcomes	Outcome Indicators	Data quality rating	End Value	Baseline	Change rating
	Count of CAPS aligned Mathematics & Language resources downloaded USE		Data not available	Data not available	
Teachers have better access to digital resources	Percentage of teachers that indicate they have access to a variety of digital content ACCESS		80% (Nov 2018)	83% (Nov 2017)	
Teachers integrate digital resources into their teaching environment	Percentage of teachers that access Professional learning communities and other collaborative platforms ACCESS		Not started		
	Percentage of teachers that upload content onto ePortal USE		3% (Nov 2018)	2% (Nov 2017)	
	Percentage of teachers that share resources with other teachers (within and outside of school) USE		75% (Nov 2018)	76% (Nov 2017)	
	eAdmin				
Principals have better access to digital school admin systems and school dashboards	Number of model schools that upload learner assessment results to a central repository ACCESS		16/16 (Term 4 2018)	16/16 (Term 1 2018)	
	Number of enhanced schools that upload learner results to a central repository ACCESS		50/50 (Term 4 2018)	50/50 (Term 1 2018)	
	Number of model schools that access learner assessment & attendance data on the dashboard ACCESS		14/16 (Term 4 2018)	11/16 (Term 1 2018)	
	Number of enhanced schools that access learner assessment & attendance data on the dashboard ACCESS		72/150 (Term 4 2018)	42/150 (Term 1 2018)	
Principals use digital school admin systems and school dashboards	Number of model and enhanced schools that use a school admin system to collect school based assessment and attendance data USE		Not started		
	Number of model and enhanced school principals that report improved data collection and management USE		Not started		

Outcomes	Outcome Indicators	Data quality rating	End Value	Baseline	Change rating
Principals support the eLearning GC	**eCulture**				
	% of school principals that express commitment to eLearning Game Changer ACCEPTANCE		66% (March 2018)	58% (March 2017)	

As argued above, a major benefit of measuring outputs separately from outcomes is that equal attention is given to the tangible (output) and intangible (outcomes) elements of an intervention. The challenges surrounding outcome measurement were discussed in Chapter 2 and in Chapter 8 guidance was provided on the conceptualisation and operationalisation of outcomes. A comparison between the eLearning output monitoring data (ref table 29) and outcome monitoring data (ref table 35) demonstrates this value add: measuring the number of teachers being trained in ICT and ICT integration is significantly easier than measuring the extent to which teachers are integrating and using ICT once back in the classroom. Similarly, tracking the number of schools that have received the Broadband and LAN roll out entails a simply "tick" on a checklist, while measuring the use of the Broadband requires for systems to be put in place and parameters around broadband use to be established. For example, does social media platforms constitute educational use? And do we only measure use during school hours? The reality is that outcome measurement requires more technical capabilities and resources, hence the tendency to focus on existing data. With this comes the risk of not paying sufficient attention to the outcomes.

The second and third gains relate to the theory of change as discussed in Chapter 6. We argued in this chapter that one of the benefits of the theory of a change is the distinction it provides between short, medium and long term outcomes. This is important as it ensures that realistic expectations are set for the achievement of such outcomes. Any assessment of outcomes must be done against the timeframe available for such outcomes to be realized. Even though the DSU was established in 2015, the eLearning Game Changer only came under the purview of the DSU towards the end of 2015. The planning stage took a further 6 months to conclude, effectively leaving three years of implementation under the Game Changer banner. In the eLearning Game Changer, the theory of change, together with the ensuing framework of "ACCESS", "ACCEPTANCE", "USE" framework presented in this chapter ensured the leadership and project teams were realistic in terms of what could be achieved within the 3-year time frame.

Table 35 demonstrates the outcomes that received the most attention during the Game Changer period by way of blue highlights. The expectation was that at a minimum the indicators relating to ACCESS to ICT and ACCEPTANCE to ICT would be expected to show positive change over the three-year period. A total of 24 short term outcome indicators were identified across the six work streams, with data ultimately being available for 20 short term indicators:

- 14 short term outcomes recorded positive changes (green rating). Teachers had increased access to wireless access points, which is mainly due to the scaled roll out of smart classroom technology. The number of CEI service desk queries related to connectivity decreased, which is a positive change as it implies that schools experience fewer connectivity problems. The CeI service desk also improved their response rates with 90% of queries being resolved within 5 working days. In the eTeachers/ eOfficials work stream the results show that the principals were more inclined to request support from district staff, with 8-percentage point improvement between March 2017 and March 2018. Where teachers and principals, are accessing the district support, they appear to be satisfied with the support received.

  Model schools, and the targeted enhanced schools were accessing the additional functionalities (data dashboards) rolled out as part of the eAdmin work stream. The eAdmin indicators related to the upload of learner assessment results showed a green score even though there was no change between the baseline measurement and the subsequent measurement. The reason for this is that the two targets were already fully achieved at baseline.

- Six of the short term indicators showed no change (orange rating). The number of academic school days affected by WAN downtime did not change much when comparing April 2018 with April 2019. Teachers' awareness of district ICT support remain unchanged, and teachers requesting support from district staff showed a slight downward trend. Access to computer laboratories was low and decreased slightly. This is not surprising given that computer laboratories accommodate very limited learner numbers. In addition, with the already too big class sizes it is not surprising that it did not present an optimal learning environment. There is also a slight downward trend in teachers' access to digital content (from 83% to 80%).

With a substantial number of short term outcomes (14 out of 20) showing a positive change, it was concluded that the eLearning Game Changer constituted a success. In the absence of separating out the short term outcomes, faulty assumptions could have been made about the accomplishments of the Game Changer given the three year implementation period.

The focus on short term outcomes did not exclude measurement of medium and long term outcomes – instead this relates to the third gain of distinguishing between output and outcome monitoring namely the ability to assess the validity of the theory of change. The theory of change captures the hypothesised causal pathways that would produce the expected outcomes. Data was thus collected for each of the outcomes in the causal pathway to affirm or reject the plausibility of each 'hypothesis' in the theory of change. In the eLearning Game Changer, eight medium-term outcome indicators (USE as per outcomes framework) were developed, with data being available and reported on for five of the indicators throughout. The long term (impact outcomes) is not reported here, as this data was only utilised as an input into the regression analysis (ref. Annexure A).

The data demonstrates the following:

- Three medium term outcomes showed a positive change (green rating). The percentage of teachers integrating ICT increased from 24% to 30% and teachers were increasingly using the wireless access points to download educational material (up from 40% to 50%). A substantial increase is noted in the percentage of model school teachers who are using a learning management system – up from 51% to 71%.
- Two medium term indicators showed no or very little change: teachers did not upload digital content onto the ePortal due to the system limitations (standing at 3% in November 2018). Teachers also did not increase their sharing of digital resources with other teachers – however this indicator was already standing at 75% at baseline measurement.

The lack of data on the short and medium term indicators can be attributed to two reasons: firstly, the systems were not in place to produce the required data: a timetable system that would record time spent in smart classrooms and computer laboratories as well as the ePortal system which experienced challenges with the tagging of resources was not in place. And secondly, the delays in rolling out the eAdmin systems meant schools did not have sufficient time to utilise the systems and therefore measuring the use of the eAdmin system was pushed back.

To arrive at a more conclusive assessment of the 'validity' of the theory of change of the eLearning intervention would require further tracking of the outcomes as well as resolving some of the data gaps over a longer period. But the basic point remains: by using the eLearning data I demonstrated how a theory based evaluation is applied in practice. By validating the theory of change, a) others in the educational field will benefit from knowing which projects and programme elements cause the desired

effects and b) whether the outcomes are indeed the "right" outcomes to achieve the ultimate impact (improved learner results).

9.6 Summary of modifications and the value add of reporting separately on outputs and outcomes in the eLearning Game Changer

Some modifications were made to Step 4a which demonstrated clear gains in monitoring the performance of the eLearning Game Changer. These modifications draw on the performance measurement and programme evaluation traditions whereby a clear distinction was made between outputs and outcome monitoring. We argued that this is a useful distinction for several reasons.

Firstly, as far as output monitoring is concerned, one is constantly checking whether implementation is happening as planned (the fidelity issue in programme monitoring); whereas with outcome monitoring, one is checking whether the outcomes are being realised as a result of successful delivery and implementation. Where outputs or deliverables are not produced as planned, this is referred to as implementation failure. This would logically lead to a response to determine why (certain) activities had not been implemented as planned and – where possible – to correct this. When outcomes are not achieved, it can be due to implementation failure or theory failure, with the latter implying that an invalid or wrong theory of change has been presumed. Why is this important? To contribute to the broader body of knowledge, one wants to be able to test programme theories of change. Being able to determine what caused the non-achievement of outcomes is therefore critical.

Secondly, tracking outputs and milestones are relatively easy, as outputs and milestones are tangible and easily quantifiable. This is not the case with outcomes. Without a dedicated focus on measuring the more 'intangible' outcomes, the tendency will be to revert mainly to output measurement.

Thirdly, by distinguishing between short-term and medium-term outcome monitoring the expectations of the implementing teams would be tempered. Not taking this into account may in fact lead to premature decisions to discontinue programmes or differences between different stakeholders emerging about the perceived success of the 'impact' of interventions.

In discussing the modifications to the Deliverology approach by way of the eLearning Game Changer data, the realities of measuring outcomes are also demonstrated. As data collection for the eLearning

Game Changer commenced, indicators were being refined, replaced and removed. On the basis of this "lived experience" of what outcome monitoring entails, it is not surprising that governments struggle to truly effect an outcomes-based approach.

The adjustments to Step 4, specifically the element related to performance monitoring suggests an expanded narrative:

From:

> Establish routines to drive and monitor performance. Once you have done some planning, you will need to know how well you are doing at implementing the plan and achieving the results promised.

To:

> Establish routines to drive and monitor performance (milestones, **outputs and outcomes).** Once you have done some planning, you will need to know how well you are doing at implementing the plan and achieving the results promised **(short vs medium term).**

Chapter 10: Conclusion and recommendations

10.1 Introduction

This thesis presented the modifications to the Deliverology approach as applied to eLearning Game Changer in the Western Cape with the aim of demonstrating that the modified approach is an effective analytical framework to assess the performance of complex social programmes.

This is, as far as I am aware, the first extensively documented case of how Deliverology was applied and tracked in a major educational intervention anywhere. In the application of the Deliverology approach I found it necessary to make certain changes and modifications to the standard Deliverology framework. This is not surprising as my overview of how Deliverology has been implemented elsewhere in the world, has shown variations in how the approach has been implemented. Our modifications to the Deliverology framework were mainly informed by certain developments in mainstream programme evaluation, such as the explicit articulation of a theory of change, the value of the logic model, greater attention to the indicator conceptualisation process and drawing a clear distinction between output and outcome monitoring.

In addressing the overall research aim two research questions were formulated:

- What are the different historical roots of Deliverology and specifically which approaches to measurement were its most influential precursors?
- How was the Deliverology approach modified in the Western Cape Government, and what gains did these modifications produce?

In the next section, I briefly summarize the main arguments and conclusions for each of the research questions before I reflect on the high-level learnings as well as recommendations that emanate from this study.

10.2 Deliverology: Its origins, influential precursors and adoption in the Western Cape Government

The first part of the thesis was devoted (a) to the history of different traditions and approaches to performance measurement in the public sector, (b) how the various performance related reforms, notably NPM, influenced performance measurement, (c) the key lessons from the history of policy

implementation and the factors that contribute to successful programmes , (d) how these different traditions culminated in Deliverology, and finally how the Deliverology approach was adopted and institutionalised in the Western Cape government in 2015.

The graph below captures best the contours of this discussion. Chapter 2 was devoted to the performance measurement approaches in public administration, new public management and new public governance. I discussed the key features of different approaches to policy implementation (including implementation research) in Chapter 3. In my final chapter in part one, I focused on Deliverology: its origins in the UK, its key features and spread across the world as well as how it was adopted by the WCG.

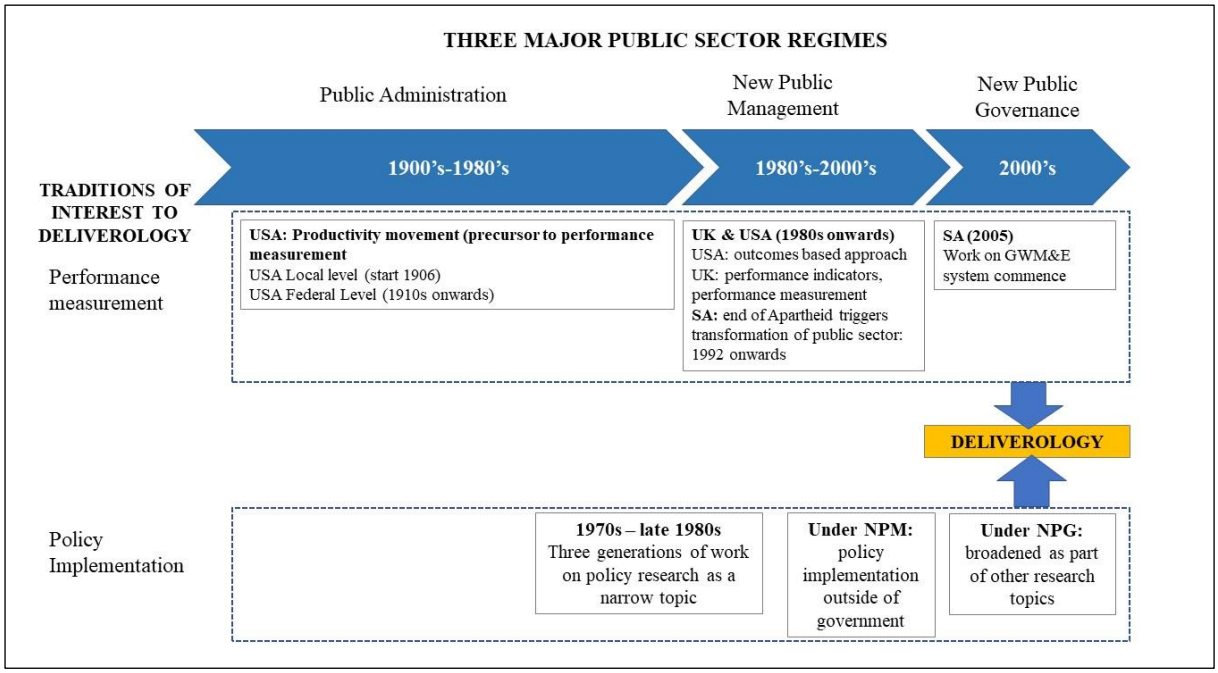


Figure 34: Historical overview of the evolution of different traditions pertaining to the performance of public sector programmes

I briefly list the salient lessons learnt from this historical overview.

10.2.1The approaches to performance measurement within the public sector and the influence of the NPM in advancing performance measurement

In Chapter 2, I showed that the earliest roots of performance measurement activities are to be found in the productivity movement of the early 1900's. Local government took the lead in attempts to curb the wide-spread corruption caused by the lack of legislative control at local level under the Jeffersonian government. The NYBMR, established in 1906, is recognised for significantly advancing the practice of performance measurement at the local level. The prevailing scientific tradition entrenched a rational

decision-making approach to the work of the bureau. This found expression in a shift from qualitative data to a quantification of social data through the NYBMR's survey activities and the development of scorecards. Additionally, the NYBMR is recognised for its ground-breaking work on budgetary reform and the development of functional budget categories to align spending to work units.

The overview shows that there are similarities as well as differences in the way performance measurement came about in the three countries covered by this study (USA, UK and South Africa). In terms of similarities, all three countries' performance related reforms were politically motivated and influenced by the current socio-economic context. Despite this similarity, there are country-level differences with the prevailing regime playing a key role.

Starting with the federal level in the USA, three stages can be distinguished during the PA regime, each reflecting a different motivation for performance measurement. The first four decades of the 20th century saw performance measurement as contributing to better government, with a strong focus on efficiency. The second stage, which started during the 1940s, was characterised by multiple budgetary reforms (for example ZBB, PPBS) to curb costs. This cost cutting rationale expanded the focus from efficiency to economy. The third stage, which started in the 1970s, had a strong political drive with politicians pushing for greater value for money in a bid to meet the needs of citizens. Although the reform efforts during this period put performance measurement on the agenda, it did not institutionalise performance measurement in the public sector.

Only during the NPM regime did the widespread adoption of performance measurement in the public sector take place. In response to the constrained environment of the early 1980s, both the UK Prime Minister Margaret Thatcher's and then USA president Ronald Reagan introduced market-related reforms with the aim of reducing the size of government and improving government efficiency. One of the reforms undertaken in the UK during this time - the FMI – called for all government units to develop performance indicators. The intent was for these performance indicators to measure outcomes, but this did not materialise, with the emphasis remaining on inputs, and to some extent outputs. The outcomes-based approach to performance measurement gained traction during President Clinton's term of office in the USA. Even though pockets of outcome measurement were being undertaken at the local level, the legislative backing for outcome measurement came by way of the CFO Act (1990) and the GPRA (1993). The NPM is not only associated with the pervasive practice of performance measurement but also the expansion of performance terminology, such as performance measurement, performance monitoring and performance indicators. Performance measurement a) must include inputs, outputs and

outcomes, b) is a continuous activity – referred to as performance monitoring and c) entails the use of performance indicators to track a set target or standard.

South Africa's performance measurement history at the national government level commenced during the NPM regime, with the post-Apartheid government (the ANC) coming into office in 1992. The ANC committed to an overhaul of the public sector, which at that stage was fragmented, authoritarian and grossly unequal in its service delivery provision. It was recognised that an integrated approach to performance management was needed as part of the transformation efforts. This resulted in the development of the GWM&E system and an outcome based approach, which drew extensively on international best practice in performance measurement.

Under the most recent regime – the NPG - a more inter-organisational way of working emerged to tackle the cross cutting, complex problems of society. The NPM resulted in a fragmented and decentralised public sector that was ill-equipped to deal with contemporary societal challenges. New governance arrangements, such as networks, had to be put in place and the centre of government had to be strengthened. Performance under the NPG becomes multi-dimensional and covers multiple levels. In addition to measuring the achievement of outcomes and process performance (for example members' commitment to goals, extent of network consultation), regime performance (for example membership of the network) also had to be measured.

Despite the recognition that governments need to focus on outcomes, my discussion of the main approaches in the history of performance measurement also reveals a clear bias towards activities and outputs that are easily measurable. It is only in more recent years that there has been a shift from outputs and deliverables to outcomes and impact in the public sector (most notably in SA through the publication of the national policy on evaluation in 2011).

10.2.2 The contribution of policy implementation and implementation research

In Chapter 3, I discussed different approaches (and generations of scholarship) in the field of policy implementation. At the overarching level, a distinction can be made between the period when policy implementation was studied within a rather narrow context of implementation research compared to when it was broadened and studied as part of other topics. In terms of the former: policy implementation research was undertaken at the intersection of public policy and public administration. Three generations of dedicated policy implementation research were undertaken, starting in the 1970s and

concluding in the 1990s: the first and second generation scholars were set on developing a theory of successful policy implementation, i.e. identifying the variables that contribute towards successful policy implementation. This resulted in the development of many frameworks that could help understand policy implementation as well as the identification of three categories of success variables namely: the policy form and content, the people involved in implementation and the organisational context and resources available for implementation.

This quest for a single theory of policy implementation was discarded by the third generation scholars, who embraced the "theoretical diversity" of the field and shifted their focus to the methodological aspects of policy implementation. The third generation scholars set out to synthesise the previous generation's work through quantitative comparative and longitudinal studies. Even though the third generation scholars are credited for their methodological contributions to implementation research, the lack of comparative, longitudinal studies resulted in little progression on the theoretical front.

With the onset of the NPG regime, policy implementation was significantly expanded. The involvement of network actors in policy implementation meant that additional factors such as network formation, network structure, the functioning of the network and the performance of the network were recognized as crucial to understand the challenges of implementation. One result was that the traditional vertically orientated and intra-organisational approach to policy implementation was replaced by a more transversal and inter-organisational ways of working.

I also discussed the overlaps between the regimes and three generations of implementation research. The importance of network actors was already being studied by first and second generation scholars some time before it gained prominence under the NPG. The involvement of external actors also features in the second generation bottom up scholars' work. These scholars called for street level bureaucrats to be part of the policy making process and recognised that as a policy moves from the macro to the micro level policy adaptation will occur due to the different role-players' interests, resource challenges and spheres of influences. Under the NPG, networks became a recognised strand of work, given the move from an intra-organisational to an inter-organisational way of working. It is understood that a variety of dynamics are at play in these network structures which affects the power dynamics. For instance, those actors who contribute more resources could ultimately "demand" greater consideration of their interests which will shape the implementation agenda. The notion of a strong centre of government comes through in the first generation scholars as well as the top down second generation theorists' work. In their view, government is responsible for setting rules and regulations, with the authoritative decision

as to what gets implemented residing with government. The NPG, in addressing the fragmentation brought about by the NPG, puts emphasis on returning control to the centre of government to enable better management of the cross cutting priorities.

10.2.3 Deliverology: origins and local adoption in the Western Cape Government

Deliverology originated in the UK in 2001 when Sir Michael Barber established the first delivery unit (the PMDU). Deliverology was initiated to address the considerable gap between policy and implementation, positing that more attention should be spent on implementation if results were to be achieved. The Deliverology approach has evolved since its inception. Not only has Barber adjusted the approach, but so have government across the globe as they have customised it to fit their context. This involved modifications to the Deliverology approach, as well as expanding the approach to include additional functions. Despite this, several success variables have been identified as different countries have experimented with the approach. The success factors include highly visible political backing, a dedicated focus on few priorities, obtaining cross cutting commitment, and instituting rigorous performance measurement routines.

The influence of the NPM and NPG, both in terms of performance measurement and policy implementation, feature strongly in Deliverology. From a NPM perspective, the principles of performance measurement, managerialism and results-based management are all evident in Deliverology. This is clearly demonstrated by the fact that Deliverology is based on the steps found in the traditional performance measurement cycle – with some nuances in place to differentiate the approach. In fact, the functions performed by a delivery unit is directly related to the steps found in the performance measurement cycle. With Deliverology commencing during the NPG regime it is not surprising that it also includes the strong focus on returning control to the centre of government. Delivery units are recognised as one of the institutional mechanisms employed by governments to achieve this objective.

In terms of the evolution of traditions in policy implementation, Deliverology bears resemblance to the second generation top down scholars who a) believed that the policy statement was the ultimate authoritative standard and b) that the policy formation phase is a distinct step from the policy implementation phase. Another point of alignment exists between Deliverology and top down second generation scholars, as well as NPG scholars: the issue of networks, and involving other actors in implementation was first raised by the bottom up second generation scholars. This work was significantly expanded under the NPG. As far as methodology is concerned, Deliverology agrees with

the three categories of success variables identified by the second generation scholars. As far as policy form and content are concerned, Deliverology commits to only a few strategic priorities which are outcome focused and supported across government. As far as the 'human' element is concerned, Deliverology emphasizes that visible political leadership is imperative, as well as the establishment of a delivery unit that "has the ear of the political leader". Strong teams also need to be put in place to ensure focused attention on the selected priorities. And finally, Deliverology require very explicit organisational arrangements and resources and that precise routines, backed by real-time data, be put in place.

Deliverology gained entry into South Africa through the National Government's introduction of an outcomes based approach in 2009. Influenced by the work of the UK PMDU, many of the Deliverology practices and principles were adopted by the newly established DPME to ensure the execution of the outcomes based approach (for example signing performance agreements with all ministers, the development of detailed delivery plans and putting in place forums to facilitate inter-departmental discussions on progress). However, it can be argued that the first full manifestation of the Deliverology approach in South Africa occurred when Premier Helen Zille established a delivery unit during her last term office in the Western Cape Province (2015). In tackling the six strategic priorities (called "Game Changers"), the DSU translated the Deliverology framework into a 4P model (prioritisation, people, performance, and problem solving). Priority-setting was done at the outset, with all the Game Changers having a clear goal statement, reflecting what needed to be achieved over the three-and-a-half year period. Dedicated Game Changer teams were established in the implementing departments as well as the DSU, supported by fully committed political and administrative leaders. To facilitate the continuous performance tracking of the Game Changer programmes, data collection systems were put in place and stocktake routines established. Delivery blockages were identified timeously, utilising the data, stocktakes and "feet on the ground" as the most common problem solving mechanisms.

10.3 Modifying and adapting the Deliverology approach to the eLearning Game Changer

Part two of the thesis provides a detailed discussion of the modification to the Deliverology approach, focusing on the gains produced by these modifications. The ability to undertake these modifications and additions was made possible through my role as a participant observer – my intimate knowledge and involvement in the performance tracking of the Game Changers enabled me to critically reflect on the Deliverology approach and to utilise my programme evaluation experience to enact certain changes. The Deliverology approach was used as the analytical framework for part two, documenting the

experiences of applying a modified Deliverology approach to the selected case of this study - the eLearning Game Changer.

10.3.1 The eLearning Game Changer as case

The eLearning Game Changer was selected as the case study for two reasons. First, it was the biggest and most systemic of all the Game Changers implemented by the DSU. The eLearning programme aimed to reach approximately 1 500 ordinary schools, 1,2 million learners and 30 000 teachers. In monetary terms, R1.1. billion (approximately US$ 73 million) was spent on the roll out of the eLearning Game Changer. Secondly, eLearning is a topic of global interest with much debate surrounding the impact of eLearning on learner performance. And most recently, the Covid-19 pandemic has also underscored the importance of eLearning, with countries needing to move their teaching and learning into a virtual environment. Educational interventions are recognised for its complexity with multiple variables coming into play as delivery is executed at various level by several agencies or stakeholders. Drawing on the eLearning Game Changer output and outcome data I demonstrate how Deliverology as a 'new' conceptual approach to public sector performance measurement in the WCG was modified over time.

10.3.2 Summary of modifications to the Deliverology approach

I have argued in this thesis that certain modifications to the Deliverology approach were required and have therefore introduced elements from programme evaluation and performance measurement to strengthen the approach.

The need for different paradigms to 'converse' with each other is not a novel idea. A good example of this is the drive for greater convergence between performance monitoring and programme evaluation in the public sector, recognising that programme evaluation can add significant value in terms of outcome measurement and explaining causal pathways (e.g. Hatry, 2013; Lahey & Nielsen, 2013; Nielsen & Hunter, 2013; Poister, 2010). Deliverology, like programme evaluation and performance measurement, aims to improve government performance. Birch and Jacob (2019), in juxtaposing Deliverology against programme evaluation reveal limited overlaps, suggesting that the two approaches might be incompatible. My case study suggests that a) there are points of overlaps between Deliverology and other paradigms and b) that the convergence of paradigms in fact strengthened the Deliverology approach.

I have also shown that it is not uncommon for delivery units across the world to modify and customise the Deliverology approach to fit the context. The DSU modifications were informed by two other long-standing performance-related traditions: programme evaluation and performance measurement. The modifications were two-fold: it entailed additions to the existing sub steps, as well as the introduction of a new sub step (Black bold font in Figure 35).

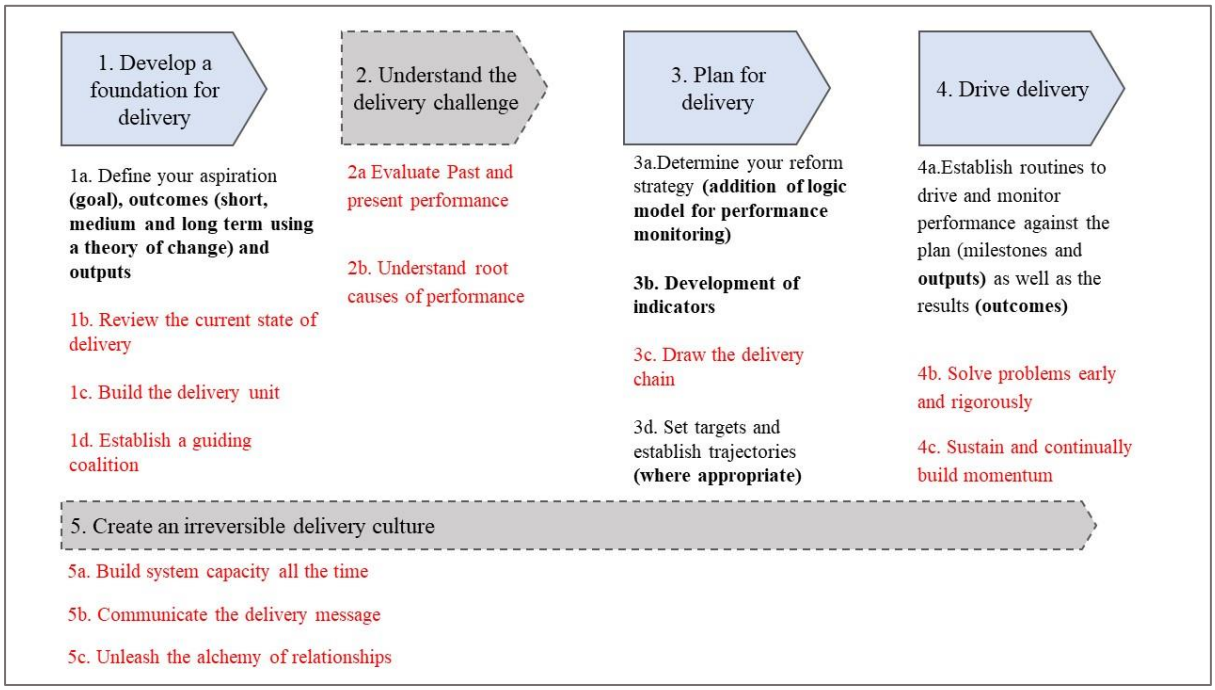


Figure 35: Additions to the Deliverology steps, and sub steps

The modifications to each of the steps were discussed in Chapters 6 to 9. Table 36 provides a summary of these modifications. Column 1 contains the original wording of the step, and relevant sub steps. Column 2 shows the points of agreement between Deliverology and programme evaluation and/ or performance measurement. Column 3 contains the gaps or shortcomings in the Deliverology framework, while column 4 describes the additions that were made to the Deliverology framework in response to these gaps. The final column contains the revised formulation of the Deliverology step and sub steps.

Table 36: Summary of modifications and additions to the Deliverology approach

Deliverology Steps and sub-steps included	Points of agreement	Gaps/ shortcomings	Additions/ Modifications	Revised step under DSU
Step 1: Develop a foundation for delivery • 1a. Define your aspiration	Utilise common terminology	Concepts not clearly defined	Importation of standard programme evaluation terminology and concepts clarified	Step 1: Develop a foundation for delivery • Define your aspiration **(goal), outcomes (short, medium and long term using a theory of change) and outputs**
	Logic matters	• Logic of programme theory not "explicit" • Lack of distinction between short, medium and long term outcomes	• Inclusion of explicit theory of change with causal pathway clearly defined • Clear distinction made between short, medium and long term outcomes • Added outputs	
	Start with aspiration			
Step 3: Plan for delivery • 3a. Reform strategies • 3c. Targets & trajectories to show estimated impact of the strategies	Theory of action is useful tool to show links between activities and outcomes	Overall coherence of delivery plan elements	Logic model developed to pull together the key elements of the intervention. Outputs included (not only milestones)	Step 3: Plan for delivery • Determine your reform strategy **(addition of logic model for performance monitoring)** • Targets & trajectories **(where appropriate)** • **Conceptualise & execute the indicators**
	Targets must be developed	Not always possible to set targets	Set targets where appropriate	
		"Metrics" covered across steps	Dedicated sub step in DSU	
		Assume mainly existing "metrics"	Expansion to include development of new indicators	
		Limited guidance on selecting indicators	Indicator selection included in the of logic model Classification of indicators made explicit	

Deliverology Steps and sub-steps included	**Points of agreement**	**Gaps/ shortcomings**	**Additions/ Modifications**	**Revised step under DSU**
Step 4: Drive Delivery • 4a. Establish routines to drive and monitor performance against the plan (milestones) as well as results	Track progress of plan and outcomes continuously	• No distinction made between output and outcome monitoring: risk – focus on easily measurable aspects • No distinction made between short and medium term outcome tracking to ensure realistic expectations with regard to what can be achieved	• Emphasis on monitoring of both output and outcomes • Distinction between short and medium term allows for formulation of more realistic expectations about programme delivery	Step 4: Drive Delivery 4a. Establish routines to drive and monitor performance against the plan (milestones and **outputs)** as well as the results **(outcomes)**

I briefly elaborate on what I believe to have been the gains of these modifications.

Define your aspiration: I concur with three aspects of step 1 in the framework: firstly, that the development of an aspiration statement should be one of the first steps in programme design; secondly, that there should be a clear alignment between such an aspiration statement and supporting "goal statements"; and finally, that a common language around performance measurement terminology must be established. At the same time, I argue that there are two remaining limitations or shortcomings in step 1. First, Deliverology does not define key terms of performance measurement adequately, and secondly, that insufficient attention is given to the important difference between short, medium- and long-term outcomes. As the DSU head of performance tracking, I thus initiated workshops (with all relevant stakeholders) on clarificative evaluation, and the development of a theory of change for the eLearning Game Changer that would eventually be accepted by everyone. The final formulation of this theory of change also included a clear distinction between short-, medium-and long-term outcomes. And finally, I decided to borrow from the standard logic model framework and included explicit references to 'outputs' in our eLearning logic model – something which is not typical practiced in the Deliverology approach.

Development of a delivery plan and reform strategies, targets and trajectories. In line with Deliverology, it is important to demonstrate the links between activities and outcomes. However, the shortcoming relates to pulling all the elements of the delivery plan together into a coherent whole, as well as ensuring the target setting is undertaken where appropriate. In addition, even though step 3 deals with targets and trajectories, the indicator selection or indicator formulation processes are addressed as part of other sub steps. There appears to be a bias towards existing metrics, yet limited guidance is provided on how to undertake this selection.

I modified this step in three ways: firstly, I utilised the logic model framework to develop our "reform strategies". The logic model is a useful tool in that it clearly demonstrates the links between outcomes and outputs, ensuring coherence of and alignment between the different elements contained in the delivery plan. The logic model also addresses the issue of target setting and the important alignment between indicators and outputs and outcomes.

The second modification consisted of the development of the indicators as part of the logic model process. This additional step ensured that the indicators are aligned to the outputs and the outcomes, preventing a focus on the easily quantifiable phenomena only. Secondly, through the process of

conceptualisation and operationalisation, outcome measurement is elucidated, showing how abstract concepts are made more measurable and finally, the systematic approach to indicator selection culminates in sound methodological decisions around appropriate data collection methods and data collection tools. The final minor modification entails an adjustment to step 3d: target setting should only be done where appropriate. In the absence of a baseline (which was the case of the eLearning Game Changer), meaningful target setting cannot be done.

Drive delivery: One of the strengths of Deliverology is its 'consistent' emphasis on tracking the delivery of an intervention. It emphasizes time and time again that outcomes need to be tracked on a continuous basis, and that data should be utilised to solve problems as they arise. I argued however, that Deliverology does not adequately distinguish between monitoring outputs from monitoring outcomes. Performance monitoring requires that both outputs and outcomes be tracked, as set out under the performance measurement tradition. The DSU therefore modified the Deliverology framework by including a clear separation of output versus outcome tracking.

Utilising the eLearning Game Changer data, I demonstrate that the first gain of monitoring outputs separately from outcomes is that one mitigates the risk of only measuring the progress of the plan (only focussing on outputs which are easily measurable) as opposed to the achievement of outcomes (which tend to be intangible and difficult to quantify). The second advantage of the clear separation between output and outcome monitoring is that it avoids having unrealistic expectations of what should be achieved within a given timeframe, which could result in the premature discontinuation of projects. Thirdly, by placing particular emphasis on outcome monitoring, it enables either the validation or rejection of the theory of change. This is not only advantageous to the officials executing the project (as it confirms whether the hypothesised causal pathway is accurate) but also advances the field of education. A validated theory of change can be used by others interested in executing eLearning programmes – either to replicate or refine further, which results in a growing body of knowledge on the critical success factors of eLearning programmes.

10.4 Key learnings and recommendations

Deliverology challenges the "business as usual" approach to government service delivery, suggesting that transformation in government can only be achieved through quality execution and boldness of reform (Barber, 2015). This is made possible through a synergistic relationship between various success variables, which is captured in the five step Deliverology framework. This study focused on the performance measurement related steps of Deliverology.

We now turn to the key learnings of this study - based on the DSU experience - to ensure optimal future performance tracking of complex interventions:

The first learning pertains to the suitability of the modified approach to the "standard" Deliverology framework. Deliverology makes it clear that it is "not M&E". Instead, the approach prescribes to a so-called 'nimble' performance measurement system that tracks a few key "metrics". Although my own background in monitoring and evaluation was indeed one of the influencing factors for modifying the Deliverology approach, the complexity of the eLearning Game Changer certainly required a more comprehensive and rigorous measurement approach. Selecting a few key indicators would not have sufficed.

Secondly, the modified approach does generate more data than one would typically encounter given the measurement of the full causal pathway (as featured in the theory of change). This required the DSU, as well as the implementing departments to make available more resource to ensure optimal use of the data. This included not only the human and budgetary resources but also a consideration of the systems and processes utilised to collect, analyse, and report on the data.

Thirdly, even though Deliverology is not viewed as a "whole of government" approach given its intense focus on a few politically motivated priorities, some broader lessons have also surfaced in presenting this modified approach to Deliverology. Through the selected case (the eLearning Game Changer) we address some of the most pertinent performance measurement challenge: we show how we have dealt with causal attribution (using a theory of change), identified the indicators (by utilising the logic model) and tackled outcome measurement (through the process of operationalisation and describing different indicator categories). This enabled sound programme design and targeted collection of granular data (at output and outcome level), which in turn improved decision making and the swift resolution of blockages.

Finally, I have demonstrated in this thesis that a modified Deliverology approach, utilising programme evaluation and performance measurement principles strengthened the measurement of a complex intervention in three ways:

- The performance measurement process was highly structured and executed against a logic model framework that was accepted by all the teams. The tabular structure assisted the teams

in identifying the links between milestones, outputs and outcomes, providing a holistic overview of all the pieces. It also helped with the distinction between outputs and outcomes, as well as the identification of suitable targets and indicators.

- The tracking of performance was rigorous and systematic because of a detailed data plan which on the whole, produced reliable granular data that could inform decision-making. Aggregate data does not enable one to pinpoint the exact challenges with delivery or whether progress is being made across the board. In the eLearning Game Changer, the collection of data at learner, teacher and school levels allowed for interventions to be customised and targeted at specific schools. The learner and teacher level data allowed us to compare the same group of teachers and learners over time, which considerably strengthened the findings.
- The performance reporting was equally structured through the stocktake routines whilst appropriate visual tools were utilised to convey the progress of the Game Changer. The distinction between outputs and outcomes, mitigated the tendency to focus on the easily measurable variables. In fact, in the last year of the Game Changer, the stocktake reports had changed significantly in emphasis towards outcome reporting.

These are important gains in a developing country context or resource-constrained situation as the modifications led to a more structured and systematic approach to performance monitoring.

I make the following recommendations as far as the modified approach is concerned:

- The modified approach to Deliverology should only be followed where it is appropriate and feasible. The study suggests that the modified approach is suited for monitoring the performance of large and complex interventions.
- Deliverology challenged the WCG's current approach to data with data systems tending to produce high level "data for monitoring" and not "data for action", i.e. granular data that is well formulated and that can support better decision making and an outcomes-based approach. This applies not only within the context of a delivery unit, but to the broader government. To change this, governments will need to embed data into the system. One option is to make data a policy or compliance requirement, such as the Canadian Government has done with all cabinet memos requiring a data plan. Another example is found in the US Federal government for education where access to state grants is linked to the provision of certain data.
- Deliverology is a resource intensive approach, and even with a "scaled down" version, additional human resources and data analytical skills are likely to be needed. Within a

developing country context, this is even more pertinent: provision for this should therefore be made in the implementing departments from the outset.

- The limited life span of the delivery unit, especially if politically motivated, should be viewed as a given. In the absence of the political leadership driving Deliverology, the onus will rest on the implementing departments to continue with the approach. Dedicated attention should be spent garnering the support of the administrative leadership to ensure longevity of the approach beyond a political term. This includes the continued tracking of performance to validate the full theory of change, especially the medium and long term outcomes.
- The modified approach increases the reliance on the delivery unit to assist with the performance measurement aspects. This could impact negatively on the irreversibility of the Deliverology approach, with the delivery unit taking on too much of the performance measurement aspects. Concerted efforts should, therefore, be placed on building the capacity of the officials and the departments involved as far as the performance measurement functions are concerned. This includes, but is not limited to, technical expertise around outcome and indicator formulation. Where possible, new data systems should be initiated, instituted, and managed by the implementing departments, not the delivery unit. If this is not possible, any data collection processes and systems managed by the delivery unit should be handed over to the implementing department(s) after a set period, with sufficient time provided for the implementing department (s) to become accustomed with the system(s).

In conclusion, in this thesis I have demonstrated the value of a highly structured approach to performance measurement (as exemplified in the elements and routines and strategies of a modified Deliverology framework) when augmented with lessons learnt from mainstream theory-based monitoring and evaluation. Given the many policy reform and service delivery challenges in South Africa – such as basic health care, food security, education, and inequality– many of our social programmes are in fact complex interventions. This study has clearly showed the value of a structured approach to tracking the performance and monitoring the outcomes of such complex programmes even within the constraints of available capacity, expertise, and governance arrangements.

Annexure A: Reporting on regression analysis of learner performance results

In the presentation to the February 2019 stocktake meeting, the statistician presented the results of a regression model that addressed the following **key question: What aspects of eLearning influence Maths performance?** Two different learner models were developed:

Table 37: Learner regression models

	Model 1a and 1b	Model 2a and b
Parameters	None – applied no pass rate	50% Pass rate
Segmentation	Primary vs High schools	Primary vs High schools
Demographic	Yes	Yes
Number of learners	17000	17000

(DSU, 2019b)

The following variables were included in the model:

Table 38: Variables included in regression model

Variables	Source	Information Level
EDUC DISTRICT	DEMOGRAPHIC	District Level
SCHOOLCATEGORY		District Level
GRADE		Learner level
GENDER		Learner level
RACE		Learner level
QUINTILE		School level
TotNbrLearners in School		School level
Attendance_R&C_YesNo	AFTERSCHOOL ATTENDANCE	Limited data
ASP_NbrAcademicPrgrms		School level
ASP_PercLregistered_Other		School level
eLL_AccessScore*	Learner Survey Data	Learner Level
eLL_Selfrating		
eLL_UseAtschoolfrq		
eLL_AttdeScr*		
eLL_Teacher_HelpUseEnc*		
eLL_StabilityScore*		
eLL_SupportScore*		

Variables	Source	Information Level
eLT_DeviceScore*	Teacher Survey Data	Teacher Level (challenges in matching teachers to learners)
eLT_CompLapTeachingScr*		
eLT_TeachSmartClass*		
eLT_ICT_IntegrtnScr*		
eLT_eContentScr*		
eLT_ TeacherRating		
eLT_ Tot_AttdeScr*		
ASS_PrincipalCmmnt*	Annual School Survey Data	School level
ASS_eContentScore*		School level
WAN_TotalMeg	WAN Data	Learner level but limited
Wan_PercEducational Content		Learner level but limited
WAN_DownTime_Minutes		Learner level
CAT_Scr	Competency Assessment Tool	Teacher Level
T4_Maths	**TARGET VARIABLE**	Learner level results
T4_PerclGT50		

(DSU, 2019b)

The results of the two different models tested were summarised as follows (presented separately for the primary and secondary (high) schools:

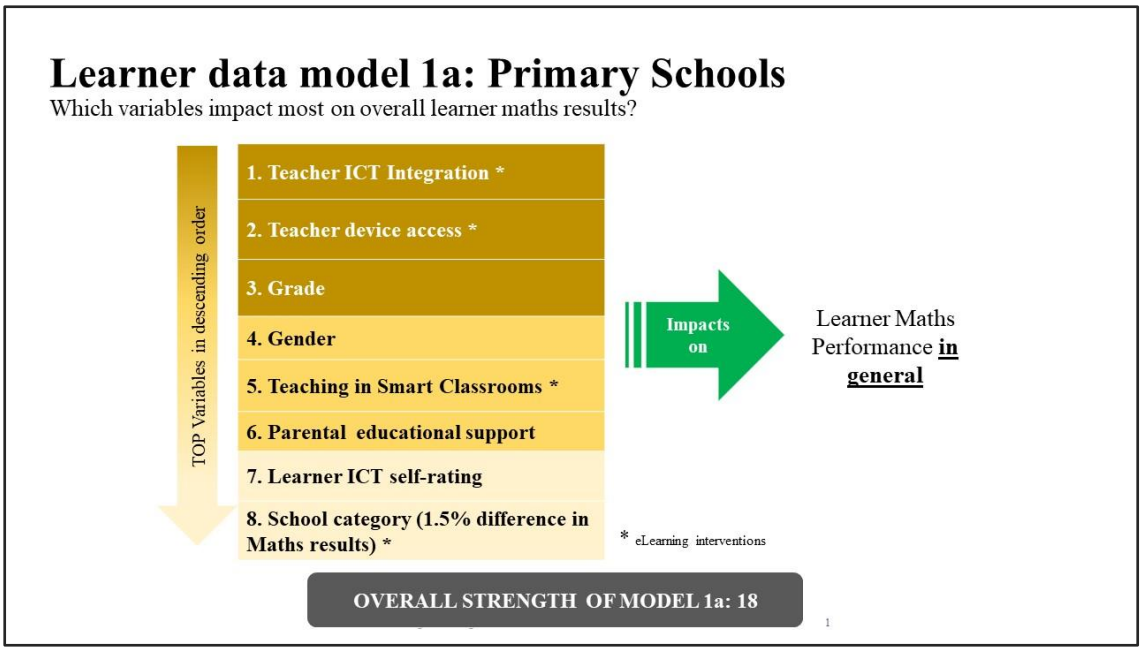


Figure 36: Summary results for model 1a: Primary schools (DSU, 2019b)

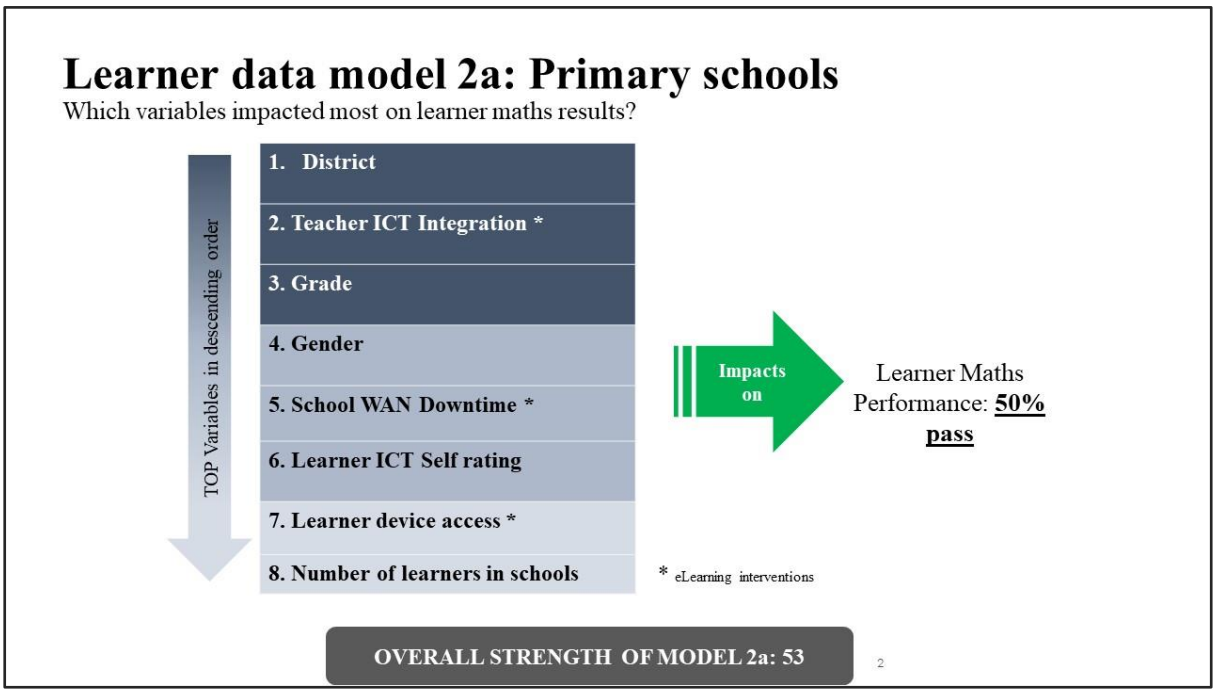


Figure 37: Summary results for model 2a: Primary schools (DSU, 2019b)

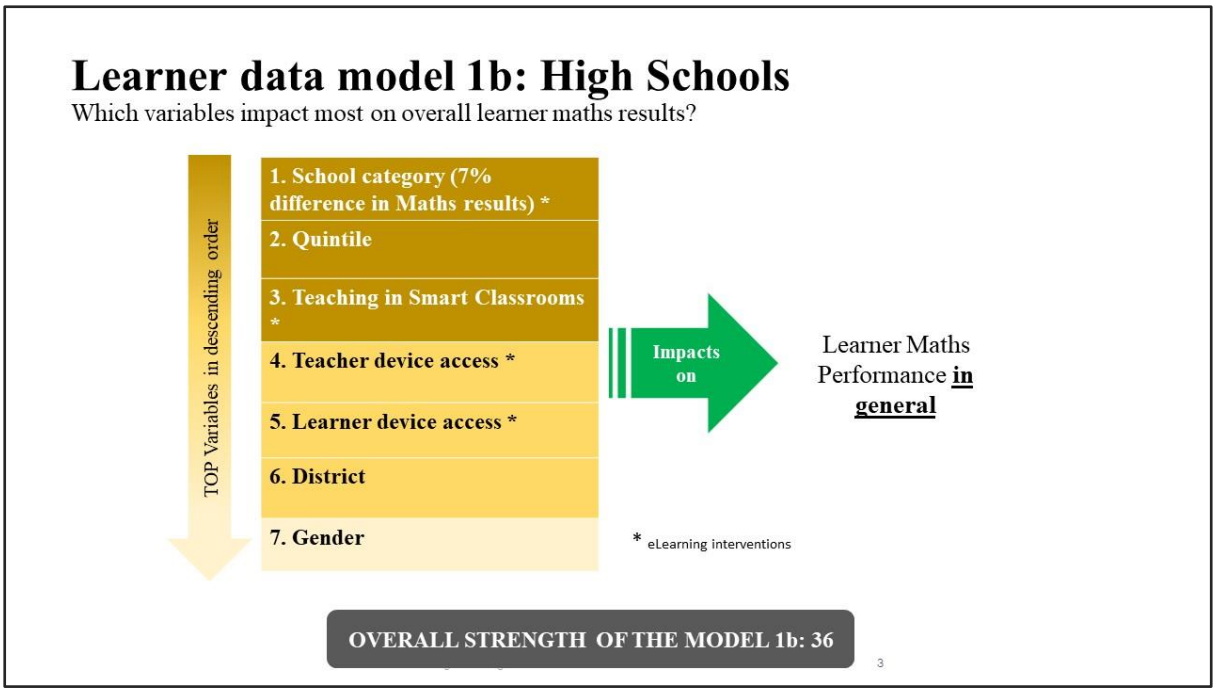


Figure 38: Summary results for model 1b: High schools (DSU, 2019b)

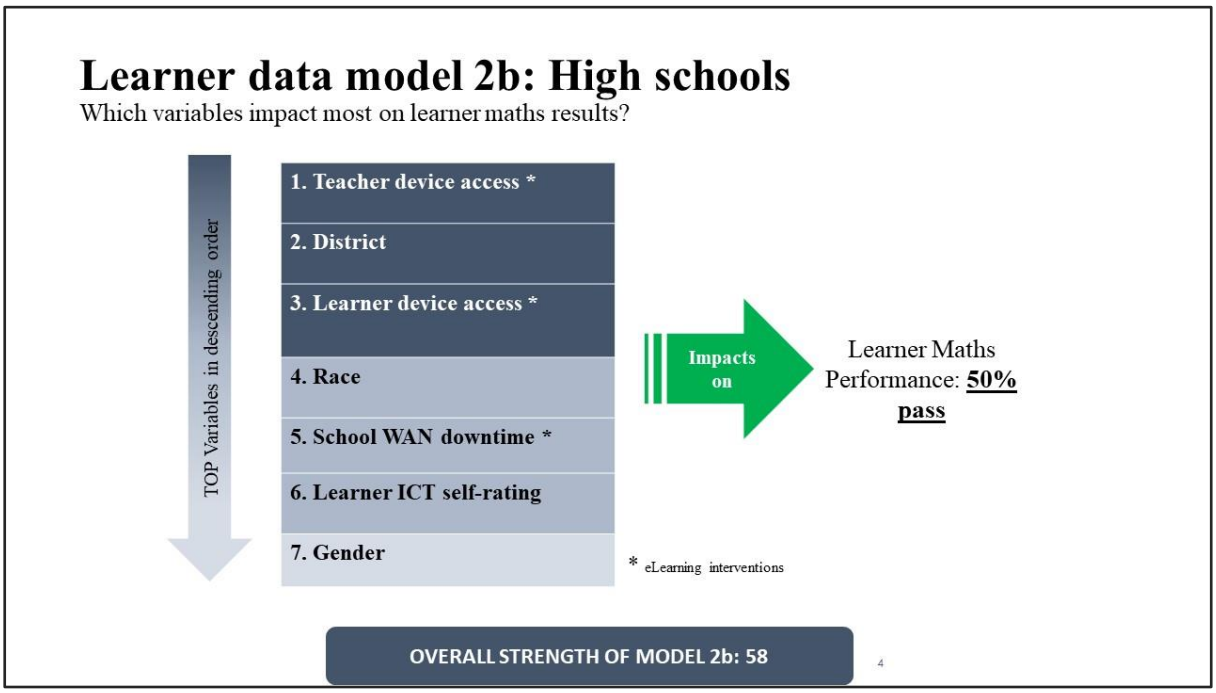


Figure 39: Summary results for model 2b: High schools (DSU, 2019b)

What do these results mean?

Results from Model 1a and 1b (Maths performance overall):

- School category (Model vs Enhanced vs Universal) features strongly in both models- much stronger in high schools
- Smart Classrooms' impact confirmed in both primary and high schools
- Currently, learner devices have higher impact in high schools, whereas teacher devices feature in both high and primary schools
- Interesting difference in variables includes parental educational support in primary and quintile in high schools

Results from Model 2a and 2b (50% pass):

- Learner devices impact more on learner performance in high schools compared to primary Schools
- WAN Downtime as well as Learners' ICT proficiency rating feature as key variables in both primary and high school models
- This is also true of districts, however we are considering the sample size of districts to understand this variable more fully
- In terms of demographic variables:
  - A learner's grade in the primary schools influences the impact eLearning has on Maths results
  - Gender features in both models but stronger in the case of primary schools
  - Race doesn't play a role in primary schools but is ranked fourth in the degree of importance in high schools

(DSU, 2019b)

Annexure B1: Data plan for performance indicators

Output statement	Performance Indicator	Indicator definition	Organisation responsible for data	Frequency of data collection	Starting date of data
eInfrastructure					
1278 Schools with WAN	Number of schools connected to Broadband (per school category)	The number of schools in each school category that have been migrated to WCG Broadband connectivity.	WCED	Ongoing	2015
366 Schools with LAN	Number of Enhanced schools with LANs	The number of schools in the enhanced school category that have an operational WCED LAN	WCED	Ongoing	2015
972 Schools with Wireless Access Points	Number of schools with wireless access points as part of the slim access point rollout (universal and enhanced schools)	The number of schools in the enhanced and universal category that had wireless access points installed as part of the WCED SLIM Lab rollout	WCED	Ongoing - Bi-Weekly - Verification	2016
School ICT queries tracked (WCG and non-WCG related)	Number of queries logged at CeI service desk (per category of query)	The number of queries logged by all public, ordinary schools and service providers via service desk, after and during school hours covering all aspects of ICT (technology, connectivity, administration systems etc)	CeI	Monthly	Apr-18
eTechnology					
1160 Schools have slim laboratories refresh	Number of schools with slim laboratory refresh (universal and enhanced schools)	The number of enhanced and universal schools where WCED have intervened to ensure a functional lab environment, either through procuring devices or leveraging existing technology	WCED	Ongoing - Bi-Weekly - Verification	2016
7530 Smart classrooms deployed with teacher devices	Number of smart classrooms (teacher devices) deployed (model and enhanced schools)	The number of technology-enabled classrooms within the category of enhanced schools and model schools. A technology enabled classroom includes as a minimum a teacher device and a projecting device	WCED	Ongoing - Bi-Weekly - Verification	2016
16183 Devices distributed to model schools	Number of learner devices that has been distributed to model schools	The number of devices distributed to model school, targeting a ratio of 1:1 but subjective to WCED budget constraints	WCED	Ongoing - Bi-Weekly - Verification	2016

Output statement	Performance Indicator	Indicator definition	Organisation responsible for data	Frequency of data collection	Starting date of data
eTeachers/ eOfficials					
32214 Principals and teachers trained in basic ICT competencies	Number of teachers trained in basic ICT competencies Number of principals trained in basic ICT competencies	The number of principals and teachers that have been trained in basic ICT competencies by the following providers: 1) CeI-Training Unit 2) Cape Teaching and Leadership Institute (CTLI) 3) WCED eLearning	WCED	Month after financial quarter	31-Oct-17
4056 Teachers trained in ICT integration	Number of teachers trained	The number of teachers that have been trained in *ICT integration by the following 3 providers 1) CeI-Training Unit 2) Cape Teaching and Leadership Institute (CTLI) 3) eLearning * ICT Integration - Professional Development that enable teachers to apply ICT in education as per the pathway	WCED	Month after financial quarter	31-Oct-17
1239 School management team members trained (including principals, deputy principals and HoDs)	Number of principals trained in ICT integration Number of deputy principals trained in ICT integration Number of HoDs trained in ICT integration	The number of school management team members (Principal, Deputy and HOD) that have been trained in *ICT integration by the following providers: 1) CeI-Training Unit 2) Cape Teaching and Leadership Institute (CTLI) 3) eLearning * ICT Integration - Training that enable school management team to apply ICT in education	WCED	Month after financial quarter	31-Oct-17
eContent					

Output statement	Performance Indicator	Indicator definition	Organisation responsible for data	Frequency of data collection	Starting date of data
Five digital resources for every topic under Maths & Language curriculum for every grade	Number of topics with digital resources available (per subject, per grade)	Total digital resources available for every topic in the CAPS curriculum for Mathematics (Pure Mathematics, Maths Literacy) and Language subjects (English/Afrikaans/IsiXhosa in Home language/First Additional language/Second additional language) in Grades 1-12.	WCED	Quarterly	2015
eAdmin					
16 model schools, 50 enhanced schools, 25 universal schools have web-based school admin systems	Number of schools with web-based school administration system in place (per school category)	The number of model, targeted enhanced and targeted universal schools where principals confirm the implementation of a school administration system	WCED	Bi-Monthly	Jan-18
16 model schools, 150 enhanced schools, have data dashboards	Number of schools where data dashboards have been made available/ rolled out (model and enhanced)	The number of model and targeted enhanced schools where principals confirm availability of a school data dashboard that displays, among others, learner enrolment, systemic test information, quarterly attendance data etc.	WCED	Bi-Monthly	Nov-18
16 model schools, 10 enhanced schools have learning management systems	Number of schools with learning management system implemented (model and enhanced)	The number of model and targeted enhanced schools where principals confirm the implementation of a learning management system (for example Moodle, Google classroom)	WCED	Bi-Monthly	Jan-18
16 Model schools, 10 Enhanced schools have parent/learner portal	Number of schools with parent/ learner portal implemented (model and enhanced)	The number of model and targeted enhanced schools where principals confirm the implementation of a learner and parent Portal (e.g. Communicator)	WCED	Bi-Monthly	Mar-18
eCulture					
Universal, enhanced and model schools sign either MOU/MOA/ Letter of Acknowledgement	Number of schools that sign MOU or MOA or Letter of Commitment	The total number of schools that have signed either a MOA (Model schools), MOU (enhanced) and Letter of Commitment (Universal) and have submitted the signed document to WCED	WCED	Ongoing	2017

Annexure B2: Data plan for outcome indicators

Intervention Level	No	Outcome	Outcome Indicator	Operational definition	Indicator calculation	Organisation responsible for data	Frequency of data collection	Starting date of data
eInfrastructure								
Teachers	1	Teachers have better connectivity	Percentage of teachers that connect to wireless access points to enable own device	The proportion of sample school teachers logging on to the WCG Broadband using the WIFI AP (access points) the past calendar month during school hours (06h00 to 18h00) in the academic school week from Monday to Friday	The number of teachers in the 111 sample schools logging on to the WCG Broadband using the WIFI AP (access points) the past calendar month during school hours (06h00 to 18h00) in the academic school week from Monday to Friday divided by the total number of teachers in the 111 sample schools and x100	CeI	Monthly	March 2018
Schools		Schools have better connectivity	Types of data being accessed and downloaded by schools	Total wireless and wired downloads per WAN category for all users during school hours (06h00 to 18h00) in the academic school week from Monday to Friday.	For 8 WAN Usage categories: wired plus wireless data downloaded by all users (learners, teachers, management) during school hours (06h00 to 18h00) the past calendar month from Monday to Friday.	CeI	Monthly	March 2018
	2		Percentage of CeI Service desk queries that relate to connectivity	The proportion of open CEI Helpdesk queries related to Broadband (problems on the internet dataline affecting connectivity to the Internet) AND open queries related to SLAN/SLIM/SIZWELAN (problems with newly installed LAN on any hardware and systems).	For 11 CeI service desk queries, aggregate for 111 sample schools the open CEI Helpdesk queries related to broadband and SLAN/ SLIM/SIZWELAN. Show as a percentage of total open queries across 11 CeI service desk categories	CeI	Monthly	March 2018

Intervention Level	No	Outcome	Outcome Indicator	Operational definition	Indicator calculation	Organisation responsible for data	Frequency of data collection	Starting date of data
			Schools data volumes	Total wired and wireless data downloaded by schools during school hours (06h00 to 18h00) in the academic school week from Monday to Friday	Wired plus wireless data downloaded in all schools during school hours (06h00 to 18h00) in the academic school week from Monday to Friday, and it includes software updates as well as educational resources. Downloaded = page views as well as downloading a resource.	CeI	Monthly	March 2018
	3		Number of academic school days affected by WAN downtime	The total number of times the 111 sample schools had no internet or network connection the past month during the academic school day, i.e. from 07h00 to 15h00.	Aggregate the academic days that the 111 sample schools had no internet or network connection the past month during the academic school day, i.e. from 07h00 to 15h00	CeI	Monthly	March 2018
	4	School have better technological support	Percentage of queries reported to CEI service desk that are resolved within five working days	Number of CEI queries, for each of the 8 categories, that are marked as "resolved" on the CEI system within 5 working days on being logged.	For each CEI Service Desk category at calendar month end: Total number of queries that are logged as "resolved" within five working days of being logged, expressed as percentage of total queries that were logged (i.e. resolved plus open queries)	A Basha	Monthly	March 2018
eTechnology								

Intervention Level	No	Outcome	Outcome Indicator	Operational definition	Indicator calculation	Organisation responsible for data	Frequency of data collection	Starting date of data
Learners	5	Learners have better access to technology	Percentage time spent in laboratory environment	Total periods per cycle taught **in a computer laboratory** divided by total periods in the cycle (Disaggregate for Maths, Maths Literacy, English First Additional Language and English Home Language)	For each teacher that submitted timesheet: Total periods per cycle taught in a Computer Lab divided by the total periods in a cycle. Aggregate up at school level and school category to derive at an average for universal and enhanced schools School cycle is typically 5 or 7 days	WCED	Annual	October 2017
	6		Percentage time spent in smart classroom	Total periods per cycle taught **in smart classroom** divided by total periods in the cycle (Disaggregate for Maths, Maths Literacy, English First Additional Language and English Home Language)	For each teacher that submitted timesheet: Total periods per cycle taught in a smart classroom divided by the total periods in a cycle. Aggregate up at school level and school category to derive at an average for universal and enhanced schools School cycle is typically 5 or 7 days	WCED	Annual	October 2017
	7		Percentage of model school learners that connect devices	The maximum number of unique learners from each model school logging on to the WCG Broadband using the wi-fi access points the past calendar month during school hours (06h00 to 18h00) in the academic school week from Monday to Friday	The day with the highest number of unique model school learners logging on to the WIFI access points the past calendar month during school hours (06h00 to 18h00) in the academic school week from Monday to Friday using their Broadband Login details are used to calculate this indicator. The total number of unique learners from 16 model schools logging on to WIFI access points are summed and then divided by the total number of model school learners. This is shown as a percentage.	CeI	Monthly	March 2018

Intervention Level	No	Outcome	Outcome Indicator	Operational definition	Indicator calculation	Organisation responsible for data	Frequency of data collection	Starting date of data
	8		Percentage of teachers teaching in computer laboratories	Number of teachers indicating they are using the computer labs to teach, irrespective of subjects	COUNT IF ticked YES option in Q13. Aggregate the yes responses and show as percentage of total teachers that completed survey	DSU	Six-monthly (repeated 3 times)	Nov 2017
	9		Percentage of teachers teaching in technology enabled classrooms	Number of teachers indicating they are using the technology enabled classrooms to teach, irrespective of subjects	COUNT IF ticked YES option in Q24. Aggregate the yes responses and show as a percentage of total teachers that completed survey	DSU	Six-monthly (repeated 3 times)	Nov 2017
eTeachers/ eOfficials								
Principals	10a	Principals are aware of the ICT integration support available to them	Percentage of principals that are aware of the ICT Integration support available to them	Percentage of principals that have contacted the district for ICT integration support	COUNT IF either selected: "I contacted district for advice on professional development" OR "I contacted the district office for advice in drafting ICT policies/plans" OR "Other District Support provided". Calculate the number of principals from public ordinary schools who selected any one of these as an option as a % of the total principals from public, ordinary schools who completed this question in the survey	WCED	Annually in March	March 2017
Teachers	10b	Teachers are aware of the ICT Integration support available to them	Percentage of teachers that are aware of the ICT Integration support available to them	Percentage of teachers that are either aware that district officials provide support or that the eLearning advisor is available to assist	COUNT IF teachers responded, "Aware of but haven't asked district for support" to either: QD3.2 "The WCED district officials provide support" or QD3.3. "The eLearning Advisor in the district is available to assist" Calculate the teachers that selected either of these options as a % of sampled teachers that completed this question in the survey	DSU	Bi-Annually	[Feb/March 2018 and repeat in October]

Intervention Level	No	Outcome	Outcome Indicator	Operational definition	Indicator calculation	Organisation responsible for data	Frequency of data collection	Starting date of data
Principals	11a	Principals recive ICT integration support	Percentage of principals that request support from District staff	Percentage of principals that have requested support from the district in either developing their teacher development plans or when drafting ICT policies or in lending other support	COUNT IF either selected: "I contacted district for advice on professional development" OR "I contacted the district office for advice in drafting ICT policies/plans" OR "Other District Support provided". Calculate the principals from public ordinary schools who selected any one of these as an option as a % of the total principals from public, ordinary schools who completed this question	WCED	Annually in March	March 2017
Teachers	11b	Teachers receive ICT integration support	Percentage of teachers that request support from District staff	Percentage of sample school teachers that indicate they are aware of, and have asked district for support in integrating ICT	COUNT IF teachers responded, "Aware of and have asked district" to either: QD3.2 "The WCED district officials provide support" or QD3.3. "The eLearning Advisor in the district is available to assist". Calculate the teachers that selected either of these options as a % of sampled teachers that completed this question in the survey	DSU	Bi-Annually	[Feb/March 2018 and repeat in October]
Principals	12a	Curriculum support officials provide better ICT support to principals	Percentage of principals that are satisfied with support they receive from respective support staff (Curriculum support officials, Professional Learning Communities	Percentage of principals that indicate their experience was either "excellent" or "good" the last time they engaged with the district about ICT and eLearning integration	Q8: Thinking of the last time you engaged with the District about ICT and eLearning integration. COUNT IF ticked either "Excellent or Good" and calculate % of principals from public, ordinary schools that have completed this question in the survey	WCED	Annually in March	March 2017

Intervention Level	No	Outcome	Outcome Indicator	Operational definition	Indicator calculation	Organisation responsible for data	Frequency of data collection	Starting date of data
Teachers	12b	Curriculum support officials provide better ICT support to teachers	Percentage of teachers that are satisfied with support they receive from respective support staff (Curriculum support officials, Professional Learning Communities	Percentage of teachers that indicate their experience was either "excellent" or "good" the last time they engaged with the district about ICT and eLearning integration	QD4: Thinking of the last time you engaged with the District about ICT and eLearning integration. COUNT teachers if ticked either "Excellent or Good" and calculate % of sample School teachers that completed this question in the survey	DSU	Bi-Annually	[Feb/March 2018 and repeat in October]
Teachers	13	Teacher use technology in their teaching practice	Percentage of teachers that use WAP to enrich educational practices	The proportion of 111 sample school teachers logging on to the WCG Broadband using the WIFI AP (access points) and accessing educational resources (websites, media streaming, social media) the past calendar month during school hours (06h00 to 18h00) in the academic school week from Monday to Friday	The number of teachers in the 111 sample schools logging on to the WCG Broadband using the WIFI AP (access points) and accessing educational resources (websites, media streaming, social media) the past calendar month during school hours (06h00 to 18h00) in the academic school week from Monday to Friday, divided by the total number of teachers in the 111 sample schools and x 100	CeI	Monthly	March 2018
	14		Percentage of teachers at model schools that use learning management systems to plan and deliver online lessons	Number of teachers at model schools, who selected the option "I use a learning management system for purpose of planning and delivering online lessons".	Number of teachers at model schools, who selected the following option in Question D1 of the teacher survey “I use ICT... learning management systems for purpose of planning and delivering online lessons". Calculate this as a percentage of total teachers that completed the survey	DSU	Bi-Annually	[Feb/March 2018 and repeat in October]

Intervention Level	No	Outcome	Outcome Indicator	Operational definition	Indicator calculation	Organisation responsible for data	Frequency of data collection	Starting date of data
	15	Teachers integrate technology into their teaching environment	Percentage of teachers that integrate ICT and e-resources in teaching practices	Percentage of teachers that integrate ICT in a) their communication with parents, fellow teachers and learners, ii) that use ICT in the classroom, iii)that use ICT for learning management systems and iv) that use ICT in learner management systems	ICT Integration score was calculated Reduced N (Teachers selecting "Never" option was excluded) Q: Please indicate in which areas you use ICT currently: 10 options (shortened): i. Send emails/ whatsapp/sms to communicate with learners ii. Send emails/ whatsapp/sms to communicate with parents iii. To communicate with fellow teachers iv. Internal admin: input on documents, timetables etc v. Learner management system: record attendance, to manage learner profile data, performance data vi. Learning management system to upload digital content vii.Learning management system for lesson planning viii. Learning management system to deliver digital content ix. Teaching and learning in classroom x. Conducting online assessments <u>The 10 options were grouped into 4 categories and weighted:</u> i) Communication items (5% weight) ii) ICT for internal administration/ learner management (15% weight) iii) ICT for learning management (digital content) (20% weight) iv) ICT for classroom teaching (60% weight)	DSU	Bi-Annually	[Feb/March 2018 and repeat in October]

Intervention Level	No	Outcome	Outcome Indicator	Operational definition	Indicator calculation	Organisation responsible for data	Frequency of data collection	Starting date of data
eContent								
Learners	16	Learners have better access to digital content	Count of CAPS aligned Mathematics & Language resources downloaded/ page views by learners by grade for each school	Total downloads and page views on average per learner accessing the ePortal	Total downloads and page views across all CAPS topics, divided by the unique number of learners that have logged in to the ePortal system	WCED	Ongoing	Nov 2017
Learners	17		Percentage of learners that indicate that a) they choose digital resources as a first choice when learning	Number of sample school learners that agree or strongly agree that digital content is more interesting than printed textbooks	Q16: Digital content is more interesting than printed textbooks. COUNT all responses ticked as STRONGLY AGREE OR AGREE and calculate as % of sampled learners across sample schools	DSU	Twice per annum	Learner Group 1: October 2017 Learner Group 2: April 2018
Teachers	18	Teachers have better access to digital resources	Percentage of teachers that indicate they have access to a variety of digital content	Number of teachers that agree or strongly agree they have access to a variety of digital content	Q41:"I have access to a variety of digital content. COUNT IT STRONGLY AGREE OR AGREE and calculate as % of sampled teachers	DSU	Bi-Annually	[Feb/March 2018 and repeat in October]
Teachers	19	Targeted teachers integrate digital resources into their teaching environment	% of Teachers that access Professional Learning Communities on e-Portal and through other collaborative platforms on a recurring basis	Number of targeted teachers that access the Moodle Professional Learning Network at least once a month	Total number of model school teachers that access the Moodle professional learning network at least once a month divided by the total number of model school teachers	WCED	Monthly	February 2019

Intervention Level	No	Outcome	Outcome Indicator	Operational definition	Indicator calculation	Organisation responsible for data	Frequency of data collection	Starting date of data
	20		Percentage of teachers that upload content onto e-Portal	Number of sample school teachers that indicate they place their content on the ePortal	Q E9: How do you share digital resources to teachers outside your school. COUNT IF ticked "I place it on the ePortal" and calculate this as a percentage of sample school teachers that indicate they use digital content (reduced N value)	DSU	Bi-Annually	[Feb/March 2018 and repeat in October]
	21		Percentage of teachers that share resources with other teachers within and outside of school	Number of sample school teachers that indicate they share resources with other teachers inside and outside of their school	Qs E7 and/or E8: Count if teacher selected "Yes, I share my resources" for question E7 and/or E8: Do you share your digital resources with other teachers in your school" "Do you share your digital resources with teachers outside of your school?". Calculate as a % of sample school teachers that completed the question	DSU	Bi-Annually	[Feb/March 2018 and repeat in October]
eAdmin								
Principals	22	Targeted principals use the digital school admin systems and school dashboards	No of model and targeted enhanced schools that upload learner assessment results to a central repository	Number of model and targeted enhanced schools that submit learner assessment data (as per learner recording and reporting sheet) on a quarterly basis	Total schools that load learner assessment data (as per learner recording and reporting sheet) on the central repository every quarter, within the stipulated timeframes	WCED	Quarterly	Oct-18
	23	Targeted principals have access to digital school admin systems and school dashboards	No of Model and targeted Enhanced schools that access learner assessment and attendance data on the dashboard	Number of model and targeted enhanced schools for which CEMIS user statistics show at least monthly viewing of learner assessment or learner attendance data as is found on CEMIS system	Number of schools for which CEMIS user statistics show at least one viewing during the month of learner assessment or learner attendance data as is found on CEMIS system	WCED	Annual	Apr-18

Intervention Level	No	Outcome	Outcome Indicator	Operational definition	Indicator calculation	Organisation responsible for data	Frequency of data collection	Starting date of data
	24	Targeted principals report better aligned data collection and management systems	No of model and targeted enhanced school principals that report improved data collection and management (school admin system and data dashboard)	Percentage of principals surveyed on their user experience of the school dashboard and school admin system that "agrees" or "strongly agrees" with the statement: a) the school admin system adds value and b) the data dashboard is user friendly	COUNT principals from public ordinary schools that selected "Agree" or "Strongly agree" to both Question 15a and 15b of Annual School Survey. Calculate this as % of all principals from public, ordinary schools that completed these questions in the Annual School survey.	WCED	Annual	Survey timing: latest Oct 2018
	25	Targeted principals use the digital school admin systems and school dashboards	No of model and targeted enhanced schools that use a school admin system to collect school based assessment and attendance data	Number of model and targeted enhanced schools that have implemented a School administration system and captures their learner assessment and attendance data on this School Administration system	Total schools that provide evidence of having a School administration system in place. The evidence must show that learner assessment and attendance data are captured on this School Administration system	WCED	Annual	Oct-18
eCULTURE								
Principals	26	School principals support the eLearning GC	Percentage of school principals that show support to eLearning Game Changer	Percentage of school principals in public ordinary schools that show support to the eLearning as indicated by the fact that they have policies, plans and people in place to implement eLearning in the school	COUNT IF indicated "YES" to Questions 1-7 and express as % of Principals that have average commitment to eLearning (31-60% of YES responses) versus a good commitment to eLearning (61% and higher). Calculate the % of those principals that have answered this question in the Annual School survey.	WCED	Annually, in March	March 2017
IMPACT indicators								
Learners		Enhanced motivation to learn	Learner Attendance	Inputted in regression analysis		WCED	Quarterly	June 2017

Intervention Level	No	Outcome	Outcome Indicator	Operational definition	Indicator calculation	Organisation responsible for data	Frequency of data collection	Starting date of data
		Improved learner outcomes	Learner results in Mathematics & Language	Inputted in regression analysis		WCED	Quarterly	June 2017
		Increased classroom participation	Percentage of learners who actively participate in classrooms	Inputted in regression analysis		WCED	Annually	2017

List of references

Aldrich, H. A. (1979). *Organisations and environments*. Englewood Cliffs, NJ: Prentice-Hall.

Alessandro, M., Lafuente, M., & Santiso, C. (2014). *Governing to deliver: Reinventing the center of government in Latin America and the Caribbean* (Monograph No. 224). Washington, DC: Inter-American Development bank.

Almquist, R., Grossi, G., Van Helden, G. J., & Reichard, C. (2012). Public sector governance and accountability. *Critical Perspectives on Accounting*, *24*(7/8), 479–487.

Anderson, J. (2010). *ICT transforming education: A regional guide*. Bangkok: UNESCO

Andrews, M., Pritchett, L., & Woolcock, M. (2015). *The challenge of building (real) state capability* (Working Paper No. 306). Cambridge: CID.

Andrews, M., Pritchett, L., & Woolcock, M. (2016). *The big stuck in state capability for policy implementation.* (Working Paper No. 318). Cambridge: CID

Auditor General South Africa. (2019). *Executive summary*. *Consolidated general report on national and provincial audit outcomes*, PFMA 2018-2019. Retrieved from https://www.agsa.co.za/Portals/0/Reports/PFMA/201819/GR/2018-19%20PFMA%20Consol%20GR.PDF

Babbie, E., & Mouton, J. (2001). *The practice of social research*. Cape Town: Wadsworth.

Barber, M. (2013). The good news from Pakistan: How a revolutionary new approach to education reform in Punjab shows the way forward for Pakistan and development aid everywhere. Retrieved from Delivery Associates https://assets.website-files.com/59ca37d5fcfbf3000197aab3/5be1df67f395d780786441d8_Pakistan%20final.pdf

Barber, M. (2015). *How to run a government: So that citizens benefit and taxpayers don't go crazy.* London: Penguin.

Barber, M. (2018). *Success delivered: How delivery units make a difference for governments and the citizens they serve*. Retrieved from Delivery Associates https://assets.website-files.com/59ca37d5fcfbf3000197aab3/5b49bdf971d09910ce35e6f7_success-delivered-delivery-associates-july2018.pdf

Barber, M., Kihn, P., & Moffit, A. (2011, Spring). Deliverology: From idea to implementation. *McKinsey on Government* [Online article]. Retrieved from McKinsey & Company https://www.mckinsey.com/industries/public-and-social-sector/our-insights/deliverology-from-idea-to-implementation

Barber, M., Rodriquez, N., & Artis, E. (2016). *Deliverology in practice: How evaluation leaders are improving student outcomes.* Thousand Oaks, CA: Corwin.

Barberis, P. (1998, Autumn). The new public management and a new accountability. *Public Administration, 76*(3), 451–470.

Bardach, E. (1977). *The implementation game: What happens after a bill becomes a law*. Cambridge, MA: The MIT Press.

Barrett, S., & Fudge, C. (Eds.). (1981). *Policy and action*. London: Methuen.

Baxter, P., & Jack, S. (2008, December). Qualitative case study methodology: study design and implementation for novice researchers. *The Qualitative Report, 13*(4), 544-559

Behn, R. (2017). How scientific is the "science of delivery"? *Canadian Public Administration, 60*(1), 89–110.

Behn, R. D. (2003). Why measure performance? Different purposes require different measures. *Public Administration Review*, *63*(5), 586–606.

Berman, E. (2002). How useful is performance measurement. *Public Performance & Management Review*, *25*(4), 348–351.

Berman, E., & Wang, X. H. (2000). Performance measurement in US counties: Capacity for reform. *Public Administration Review*, *60*(5), 409–420.

Berman, P. (1978). The study of macro-and micro-implementation. *Public Policy*, *26*(2), 157–184.

Bevan, G., & Hood, C. (2006). What's measured is what matters: Targets and gaming in the English public health care system. *Public Administration*, *84*(3), 517–538.

Bhuasiri, W., Xaymoungkhoun, O., Zo, H., Rho, J. J., & Ciganek, A. P. (2012). Critical success factors for eLearning in developing countries: A comparative analysis between ICT experts and faculty. *Computers and Education*, *58*(2), 843–855.

Birch, L., & Jacob, S. (2019). "Deliverology" and evaluation: A tale of two worlds. *Canadian Journal of Program Evaluation*, *34*(2), 303–328.

Blair, R. (2000). Policy implementation networks: The impact of economic development on new public management. *International Journal of Economic Development*, *2*(4), 511–536.

Blalock, A. B. (1999). Evaluation research and the performance management movement: From estrangement to useful integration? *Evaluation, 5*(2), 117–149.

Blau, P. M. (1982). Structural sociology and network analysis: An overview. In P. V. Marsden & N. Lin (Eds.), *Social structure and network analysis* (pp. 273–279). Beverly Hills, CA: Sage.

Boland, T., & Fowler, A. (2000). A systems perspective of performance management in public sector organisations. *International Journal of Public Sector Management*, *13*(5), 417-446

Bouckaert, G. (1990). The history of the productivity movement. *Public Productivity & Management Review*, *14*(1), 53–89.

Bouckaert, G. (1992). Public productivity in retrospective. In M. Holzer (Ed.), *Public productivity handbook* (pp. 15–46). New York, NY: Marcel Dekker.

Bouckaert, G., & Halligan, J. (2006). *A framework for comparative analysis of performance management.* Paper presented to the Study Group on Productivity and Quality in the Public Sector, Conference of European Group of Public Administration, Universita Bocconi, Milan.

Bovaird, T., & Loffler, E. (2003). Understanding public management and governance. In T. Bovaird & E. Loffler (Eds.), *Public management and governance* (pp. 3–14). (2nd ed.). Abingdon: Routledge.

Bovens, M. (2006). *Analysing and assessing public accountability: A conceptual framework*. European Governance Papers (No. C-06-01). Retrieved from http://www.connex-network.org/eurogov/pdf/egp-connex-C-06-01.pdf

Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. *European Law Journal*, *13*(4), 447–468.

Bovens, M. (2010). Two concepts of accountability: Accountability as a virtue and as a mechanism. *West European Politics*, *33*(5), 946–967.

Bovens, M., Schillemans, T., & Hart, P. T. (2008). Does public accountability work? An assessment tool. *Public Administration*, *86*(1), 225–242.

Burstein, C., & Sedlak, K. (1998). The federal productivity improvement effort: Current status and future agenda. *National Productivity Review*, *7*(2), 122–134.

Cameron, R. (2009). New public management reforms in the South African public service: 1999–2009. *Journal of Public Administration*, *44*, Special issue 1, 910–942.

Cameron, R., & Tapscott, C. (2000). The challenges of state transformation in South Africa. *Public Administration and Development*, *20*(2), 81–86.

Cargill, J. C. (2018). *Overview of DSU and deliverology*. PowerPoint presentation presented at Cabinet Bosberaad, Cape Town.

Cargill, J. C., & Mouton, C. (2018). *Public sector delivery approaches & delivery units: International visits to delivery units; results/data focused entities; and public sector performance experts.* Unpublished report, Delivery Support Unit, Western Cape Government, Cape Town.

Carley, M. (1981). *Social Measurement and social indicators. Issues of policy and theory*. London: George Allen & Unwin

Carter, N. (1991, Spring). Learning to measure performance: The use of indicators in organizations. *Public Administration*, *69* (1), 85–101.

Carter, N., Klein, R., & Day, P. (1992). *How organisations measure success: The use of performance indicators in government*. London: Routledge.

Chatterji, M. (2016). Causal inferences on the effectiveness of complex social programs: navigating assumptions, sources of complexity and evaluation design challenges. *Evaluation and Program planning, 59,* 128-140

Chen, H.T. (n.d.). *Theory-driven evaluation: conceptual framework, methodology and application.* [Powerpoint presentation]. Retrieved from https://ceval.de/modx/uploads/pdf/Chen_presentation.pdf

Chen, H. T. (2005). *Practical program evaluation: Assessing and improving planning, implementation, and effectiveness*. (1st ed.). Thousand Oaks, CA: Sage.

Chenhall, R. H., & Langfield-Smith, K. (2007). Multiple perspectives of performance measures. *European Management Journal*, *25*(4), 266–282.

Christensen, T. (2012). Post-NPM and changing public governance. *Meiji Journal of Political Science and Economics*, *1*, 1–11.

CLEAR. (2012). *African monitoring and evaluation systems: Exploratory case studies*. *A collection of case studies facilitated by the CLEAR Initiative – WITS.* Retrieved from https://www.theclearinitiative.org/sites/clearinitiative/files/2016-04/african_M%26E_cases_1.pdf

Cloete, F. (2000). *At full speed the tiger cubs stumbled: Lessons from South East Asia about sustainable public service delivery*. Pretoria: HSRC Press.

Cloete, F. (2009). Evidence-based policy analysis in South Africa: Critical assessment of the emerging government-wide monitoring and evaluation system. *South African Journal of Public Administration*, *44*(2), 293–311.

Centre for Research on Evaluation, Science and Technology (CREST). (N.d.) *PGD in Monitoring and Evaluation methods. General principles and paradigms in evaluation studies* [Module text]. Stellenbosch: Stellenbosch University, 10225-772

Cook, T. J., Vansant, J., Stewart, L., & Adrian, J. (1995). Performance measurement: Lessons learned for development management. *World Development*, *23*(8), 1303–1315.

Coryn, C.L.S., Noakes, L.A., Westine, C.D., & Schroter, D.C. (2011). A systematic review of theory-driven evaluation practice from 1990 to 2009. *American Journal of Evaluation, 32(*2), 199-226

Cox, R. W. (2002). Performance measurement: Cultural and social-psychological influences. In K. van der Molen, A. van Rooyen, & B. van Wyk (Eds.), *Outcomes-based governance: Assessing the results* (pp. 160–185). Sandown: Heinemann.

Cronjé, F. (2010, November 19). Policies need to be implemented. *Business Day*, p. 14.

Cross, M., Adam, F., & Ojo, E. (2009). *Preparing educators for an ICT-mediated teaching environment in South African high education: Imperatives, options and choices*. Paper presented to the World Conference on ELearning in Corporate, Government, Healthcare, and Higher Education (ELEARN), Vancouver, Canada.

DeGroff, A., & Cargo, M. (2009). Policy implementation: Implications for evaluation. *New Directions for Evaluation*, *124*, 47–60.

Department of Education. (2004). *White paper on e-Education: Transforming learning and teaching through information and communication technologies* (ICTs.) Pretoria: Government Printer.

Department of Performance Monitoring and Evaluation. (2011). *National evaluation policy framework*. Retrieved from https://www.dpme.gov.za/publications/Policy%20Framework/National%20Evaluation%20Policy%20Framework.pdf

Department of Performance Monitoring and Evaluation. (N.d.a). *Operation Phakisa*. Retrieved from https://www.operationphakisa.gov.za/Pages/Home.aspx

De Vries, M., & Nemec, J. (2013). Public sector reform: An overview of recent literature and research on NPM and alternative paths. *International Journal of Public Sector Management*, *26*(1), 4–16.

Dhillon, L., & Vaca, S. (2018). Refining theories of change. *Journal of MultiDisciplinary Evaluation*, *14*(3), 64–87.

Downs, G. W., & Larkey, P. D. (1986). *The search for government efficiency: From hubris to* helplessness. Philadelphia, PA: Temple University Press.

Dubnick, M. (2005). Accountability and the promise of performance: In search of the mechanisms. *Public Performance & Management Review*, *28*(3), 376–417.

Duignan, P. (2003, Autumn). Mainstreaming evaluation or building evaluation capability? Three key elements. *New Directions for Evaluation*, 99), 7–21.

Du Toit, I. (2005). *The effect of ICT curriculum support on the measured skills levels of learners of two sub-projects of the Khanya Project*. Paper presented to the 8th IFIP World Conference on Computers in Education (WCCE), Stellenbosch.

Ehrler, F. (2012). New public governance and activation. *International Journal of Sociology and Social Policy*, *32*(5/6), 327–339.

English, B.J, Cummings, R., & Straton, R. G. (2002). Choosing an evaluation model for community crime prevention programs. *Crime Prevention Studies*, *14*, 119-170.

Erridge, E. (2003). Contracting for public sector organisations. In T. Bovaird & E. Loffler (Eds.), *Public management and governance* (pp. 95–108). Abingdon: Routledge.

Funnel, S.C & Rogers, P.J. (2011). *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models.* San Francisco (CA): Jossey-Bass

Friedman, J. (2011). Sticking to the numbers: performance monitoring in South Africa, 2009-2011. Innovations for successful societies. Princeton University. Retrieved from https://successfulsocieties.princeton.edu/sites/successfulsocieties/files/Sticking%20to%20the%20numbers_1.pdf

Fryer, K., Antony, J., & Ogden, S. (2009). Performance management in the public sector. *International Journal of Public Sector Management*, *22*(6), 478–498.

Gauteng Provincial Treasury. (2019). Vote 1 – Office of the Premier. Retrieved from http://www.treasury.gov.za/documents/provincial%20budget/2019/3.%20Estimates%20of%20Prov%20Rev%20and%20Exp/GT/2.%20Estimates%20of%20Prov%20Rev%20and%20Exp/GT%20-%20Vote%2001%20-%20Office%20of%20the%20Premier.pdf

General Accounting Office. (1993). *Performance budgeting: State experiences and implications for the federal government*. United States General Accounting Office (No GAO/AFMD-93-41). Retrieved from https://www.gao.gov/assets/220/217424.pdf

Gilder, G. (1975). Public productivity review. *Public Productivity Review*, *1*(1), 4–8.

Goggin, M. L. (1986). The "too few cases / too many variables" problem in implementation research. *The Western Political Quarterly*, *39*(2), 328–347.

Goggin, M. L., Bowman, A. O., Lester, J. P., & O'Toole, L. (1990). *Implementation theory and practice: Towards a third generation*. Glenview, IL: Scott Foresman.

Gold, J. (2014). *International delivery: Centres of government and the drive for better policy implementation*. (MOWAT Research No. 96). Retrieved from

https://www.instituteforgovernment.org.uk/sites/default/files/publications/International%20Delivery%20report.pdf

Gold, J. (2017). *Tracking delivery: Global trends and warning signs in delivery units*. Retrieved from https://www.instituteforgovernment.org.uk/sites/default/files/publications/Global%20Delivery%20report.pdf

Golder, G. (1975). Public sector productivity. *Public Productivity Review*, *1*(1), 4–8.

Goldman, I., Byamugisha, A., Gounou, A., Smith, L. R., Ntakumba, S., Lubanga, T., Damase, S., & Rot-Munstermann, K. (2018). The emergence of government evaluation systems in Africa: The case of Benin, Uganda and South Africa. *African Evaluation Journal*, *6*(1), 1–11.

Goldman, I., Mathe, J. E., Jacob, C., Hercules, A., Amisi, M., Buthelezi, T., & Sadan, M. (2015). Developing South Africa's national evaluation policy and system: First lessons learned. *African Evaluation Journal*, *3*(1), 1–9.

Goldman, I., Engela, R., Akhalwaya, Il., Gasa, N, Leon, B., Mohamed, H., Phillips, S. (2012, September). *Establishing a national M&E system in South Africa*. Special Series on the Nuts & Bolts of M&E Systems (No.21). Washington D.C.: World Bank

Gorard, S. (2010). Measuring is more than assigning numbers. In G. Walford, E. Tucker & M. Viswanathan (Eds.), *The SAGE Handbook of measurement* (pp. 389-408). Los Angeles: Sage.

Gray, A., & Jenkins, B. (1982, Winter). Policy analysis in British central government: The experience of par. *Public Administration*, *60*, 429–450.

Gray, A., Jenkins, B., Flynn, A., & Rutherford, B. (1991). The management of change in Whitehall: The experience of the FMI. *Public Administration*, *69*(1), 41–59.

Gray, A., & Jenkins, W.I. (1986). Accountable management in British central government: Some reflections on the Financial Management Initiative. *Financial Accountability & Management*, *2*(3), 171–186.

Gray, A., & Jenkins, W. I. (1985). *Administrative politics in British government.* Brighton: Wheatsheaf Books.

Gulf Times. (2020, January 10). South Africa business confidence at 3-decade low in 2019, p. 2.

Guthrie, J., & English, L. (1997). Performance information and programme evaluation in the Australian public sector. *The International Journal of Public Sector Management, 10* (3) 154–164.

Guy Peters, B., & Pierre, J. (2001). Developments in intergovernmental relations: Towards multi-level governance. *Policy and Politics*, *29*(2), 131–135.

Halachmi, A. (2002a). Performance measurement, accountability, and improved performance. *Public Performance & Management Review*, *25*(4), 370–374.

Halachmi, A. (2002b). Performance measurement and government productivity. *Work Study*, *51*(2), 63–73. doi:10.1108/00438020210418782

Halachmi, A. (2002c). Who gets what, when, and how: Performance measures for accountability? For improved performance? *International Review of Public Administration*, *7*(1), 85–95.

Halachmi, A. (2005). Performance measurement is only one way of managing performance. *International Journal of Productivity and Performance Management*, *54*(7), 502–516.

Halligan, J. (2007). Accountability in Australia: Control, paradox, and complexity. *Public Administration Quarterly*, *31*(4), 453–479.

Hanf, K., & Scharpf, F. W. (Eds.). (1978). *Interorganizational policy making: Limits to coordination and central control.* London: Sage.

Hargrove, E. C. (1975). *The missing link: The study of the implementation of social policy*. Washington D.C.: Urban Institute.

Harrison, H., Birks, M., Franklin, R., & Mills, J. (2017). Case study research: foundations and methodological orientations. *Forum: Qualitative Social Research*, *18*(1),

Hatry, H., & Fisk, D. (1971). *Improving productivity and productivity measurement in local government*. Washington, DC: Urban Institute.

Hatry, H., and others. (1974). *Measuring the effectiveness of basic municipal services*. Washington, DC: Urban Institute.

Hatry, H. P. (2002). Performance measurement: Fashions and fallacies. *Public Performance & Management Review*, *25*(4), 352-358.

Hatry, H. P. (2013). Sorting the relationships among performance measurement, program evaluation, and performance management. In S. B. Nielsen & D. E. K. Hunter (Eds.), *Performance management and evaluation.* New Directions for Evaluation, 137, 19–32.

Head, B. W. (2018). Forty years of wicked problems literature: Forging closer links to policy studies. *Policy and Society*, *38*(2), 180–197.

Heinrich, C. J. (2003). Measuring public sector performance and effectiveness. In B. Guy Peters & J. Pierre (Eds.), *Handbook of public administration* (pp. 25–37). London: Sage.

Hill, M., & Hupe, P. (2003). The multi-layer problem in implementation research. *Public Management Review*, *5*(4), 471–490.

Hill, M., & Hupe, P. (2014). *Implementing public policy* (3rd ed.). London: Sage.

Hjern,B. (1982, August). Implementation research: the link gone missing. *Journal of Public Policy, 2*(3), 301-308

Hjern, B. & Hull, C. (1982). Implementation research as empirical constitutionalism. In B. Hjern & C. Hull (Eds.), *Implementation beyond hierarchy*, Special issue of the European Journal of Political research.

Hjern, B. & Porter, D.O. (1981). Implementation structures: A new unit of administrative analysis. *Organization studies, 2*(3), 211-227

Ho, A. (2007). GPRA after a decade: Lessons from the Government Performance and Results Act and related federal reforms: Introduction. *Public Performance & Management Review*, *30*(3), 307–311.

Hodges, R. (2012). Joined-up government and the challenges to accounting and accountability researchers. *Financial Accountability & Management*, *28*(1), 26–51.

Hogwood, B. W., & Gunn, L. (1984). *Policy analysis for the real world*. London: Oxford University Press.

Holzer, M., & Callahan, K. (1998). *Government at work: Best practices and model programs*. Thousand Oaks, CA: Sage.

Hood, C. (1995). The "new public management" in the 1980s: Variations on a theme. *Accounting, Organizations and Society*, *20*(2/3), 93–109.

Hood, C. (2006, July/August). Gaming in Targetworld: The targets approach to managing British public services. *Public Administration Review*, 515–521.

Hood, C., James, O., Jones, G., Scott, C., & Travers, T. (1998, April–June). Regulation inside government: Where new public management meets the audit explosion regulation. *Public Money and Management*, 61–68.

Hope, K. R. (2001). The new public management: Context and practice in Africa. *International Public Management Journal*, *4*(2), 119–134.

Howie, S. J. (2010). ICT-supported pedagogical policies and practices in South Africa and Chile: Emerging economies and realities. *Journal of Computer Assisted Learning*, *26*(6), 507–522.

Humphrey, C., Miller, P., & Scapens, R. W. (1993). Accountability and accountable management in the UK public sector. *Accounting, Auditing & Accountability Journal*, *6*(3), 7–29.

Hupe, P., & Sætren, H. (2015). Comparative implementation research: Directions and dualities. *Journal of Comparative Policy Analysis: Research and Practice*, *17*(2), 93–102.

Hupe, P. L. (2011). The thesis of incongruent implementation: Revisiting Pressman and Wildavsky. *Public Policy and Administration*, *26*(1), 63–80.

Hyndman, N. S., & Anderson, R. (1998). Performance information, accountability and executive agencies. *Public Money and Management*, *18*(3), 23–30.

Istrate, O. (2013). Criteria for elearning programmes evaluation. *Journal of Advanced Distributed Learning Technology*, *1*(1), 27–35.

Jakobsen, M. L., Baekgaard, M., Moynihan, D. P., & Van Loon, N. (2017). Making sense of performance regimes: Rebalancing external accountability and internal learning. *Perspectives on Public Management and Governance*, *1*(2), 127–141.

Johansson, R. (2003). Key note speech at the international conference *"Methodologies in Housing Research,"* Royal Institute of Technology in cooperation with the International Association of People–Environment Studies, Stockholm, September 22-24, 2003. Retrieved from http://www.psyking.net/htmlobj-3839/case_study_methodology-_rolf_johansson_ver_2.pdf

Jordan, M. M., & Hackbart, M. M. (1999, Spring). Performance budgeting and performance funding in the States: A status assessment. *Public Budgeting & Finance*, *19* (1) 68–88.

Joyce, P. G. (1993, Winter). Using performance measures for federal budgeting: Proposals and prospects. *Public Budgeting & Finance*, *13* (4) 3–17.

Joyce, P. G. (1997, Fall). Using performance measures for budgeting: A new beat, or is it the same old tune? *New Directions for Evaluation*, *75*, 45–61.

Kaplan, R. S., & Johnson, H. T. (1987). *Relevance lost: The rise and fall of management accounting*. Boston, MA: Harvard Business School Press.

Kaplan, R. S., & Norton, D. P. (1992, January–February). The balanced scorecard: Measures that drive performance. *Harvard Business Review*, *70*(1), 71–79.

Kawulich, B.B. (2005, May). Participant Observation as a data collection method. *Forum: Qualitative Social Research*, *6*(2), Art. 43

Kearns, K. P. (1994, March - April). The strategic management of accountability in nonprofit organizations: An analytical framework. *Public Administration Review*, *54*(2),185–192.

Keast, R., & Mandell, M. P. (2013). Network performance: A complex interplay of form and action. *International Review of Public Administration*, *18*(2), 27–45.

Keegan, D. P., Eiler, R. G., & Jones, C. R. (1989). Are your performance measures obsolete? *Management Accounting*, *70*, 45–50.

Kennerley, M., & Neely, A. (2002). A framework of the factors affecting the evolution of performance management systems. *International Journal of Operations & Production Management*, *22*(11), 1222–1239.

Koma, S., & Tshiyoyo, M. (2015). Improving public service delivery in South Africa: A case of administrative reform. *African Journal of Public Affairs*, *8*(2), 30–42.

Koppenjan, J., Karre, P. M., & Termeer, K. (Eds.). (2019). *Smart hybridity: Potential and Challenges of New Governance Arrangements*. The Hague: Eleven International Publishing.

Koppenjan, J. & Klijn, E.H. (2004). *Managing uncertainties in networks: A network approach to problem solving and decision making*. London: Routledge.

Koppenjan, J. & Koliba, C. (2013). Transformations towards new public governance: Can the new paradigm handle complexity? *International Review of Public Administration*, *18*(2), 1–8.

Lafuente, M., & González, S. (2018). *Do delivery units deliver? Assessing government innovation*. (Technical note no. 1431). Washington, D.C.: Inter-American Development Bank

Lahey, R., & Nielsen, S. B. (2013). Rethinking the relationship among monitoring, evaluation, and results-based management: Observations from Canada. *New Directions for Evaluation*, *137*, 45–56.

Lane, J.-E. (2000). *New public management: An introduction*. London: Routledge.

Lapsley, I. (2008). The NPM agenda: Back to the future. *Financial Accountability & Management*, *24*(1), 77–96.

Larbi, G. A. (1999, September). *The new public management approach and crisis states for social development*. (Discussion Paper No.112). Geneva: United Nations Research Institute for Social Development

Laumann, E. O., & Knoke, D. (1987). *The organisational state: Social choice in national policy domains.* Madison, WI: Wisconsin University Press.

Lauth, T. P. (1985, Spring). Performance evaluation in the Georgia budgetary process. *Public Budgeting & Finance*, 67–82.

Lewis, J. M., & Triantafillou, P. (2012). From performance measurement to learning: A new source of government overload? *International Review of Administrative Sciences*, *78*(4), 597–614.

Lindberg, S. I. (2013). Mapping accountability: Core concept and subtypes. *International Review of Administrative Sciences, 79*(2), 202–226.

Lipsky,M. (1980). *Street-level bureaucracy: dilemmas of the individual in public services*. New York, N.Y.: Sage.

Loffler, E. (2003). Public governance in a network society. In T. Bovaird & E. Loffler (Eds.), *Public management and governance* (pp.215–232). Abingdon: Routledge.

Lynch, T., & Day, S. E. (1996). Public sector performance measurement. *Public Administration Quarterly*, *19*(4), 404–419.

Lynn, L. E. (2010). What endures? Public governance and cycle of reform. In S. P. Osborne (Ed.), *The new public governance? Emerging perspectives on the theory and practice of public governance* (pp. 105–122). London: Routledge.

Mabelebele, J. M. (2006). Prospects and challenges of implementing projects in public service, South Africa. In *Proceedings of the 2006 PMSA International Conference: Growth and Collaboration for a Project Management Profession* (pp. 247–254).

Majone, G., & Wildavsky, A. (1978). Implementation as evolution. In H. Freeman (Ed.), *Policy studies review annual* (pp. 103-117). Beverly Hills, CA: Sage.

Manning, N., & Watkins, J. (2013). *GET note: Targeting results, diagnosing the means: Innovative approaches for improving public sector delivery* (No. 98815). Retrieved from http://documents1.worldbank.org/curated/en/136171467986259116/pdf/98815-BRI-VC-ADD-SERIES-PUBLIC-Box393182B.pdf

Marr, B. (2009). *Managing and delivering performance: How government, public sector and not-for-profit organizations can measure and manage what really matter*. London: Elsevier.

Matheson, A. (2002). Outcome-focused management in OECD countries. In K. van der Molen, A. van Rooyen, & B. van Wyk (Eds.), *Outcomes-based governance: Assessing the results* (pp. 7–18). Sandown: Heinemann.

Matland, R. E. (1995). Synthesizing the implementation literature: The ambiguity-conflict model of policy implementation. *Journal of Public Administration Research and Theory*, *5*(2), 145–174.

Mayne, J. (2007). Challenges and lessons in implementing results-based management. *Evaluation, 13*(1):87–109.

Mazmanian, D. A., & Sabatier, P. A. (1989). *Implementation and public policy*. Lanham, MD: University Press of America.

McLaughlin, J.A. & Jordan, G.B. (2010). Using Logic models. In J.S. Wholey, H.P. Hatry, & K.E. Newcomer (Eds.), *Handbook of practical program evaluation* (3rd ed.) (pp.100-124). San Francisco, CA: Jossey-Bass.

McCourt, W. (2012). Can top-down and bottom-up be reconciled? Electoral competition and service delivery in Malaysia. *World Development*, *40*(11), 2329–2341.

McDavid, J. C., Huse, I., & Hawthorn, L. R. L. (2013). *Program evaluation and performance measurement: An introduction to practice* (2nd e-d.). Los Angeles, CA: Sage.

McNab, R. M. & Melese, F. (2003, Summer). Implementing the GPRA: Examining the prospects for performance budgeting in the federal government. *Public Budgeting & Finance, 23* (2), 73-95

McQuaid, R. W. (2010). Theory of organizational partnerships: Partnership advantages, disadvantages and success factors. In S. P. Osborne (Ed.), *The new public governance? Emerging Perspectives on the Theory and Practice of Public Governance* (pp. 127–148). London: Routledge.

Melkers, J., & Willoughby, K. (1998). The state of the states: Performance-based budgeting requirements in 47 out of 50. *Public Administration Quarterly*, *58*(1), 66–73.

Merriam, S. B. (1988*). Case study research in education: A qualitative approach*. San Francisco, CA: Jossey-Bass.

Milakovich, M. E. (1992). Total quality management for public service productivity improvement. In M. Holzer (Ed.), *Public productivity handbook* (pp. 577–603). New York, N.Y.: Marcel Dekker.

Miller, K. (2005). *Public sector reform: Governance in South Africa*. Hants: Ashgate.

Mimba, N. P. S. H., Van Helden, J., & Tillema, S. (2007). Public sector performance measurement in developing countries: A literature review and research agenda. *Journal of Accounting & Organizational Change*, *3*(3), 192–208.

Minnaar, F. (2006). Integrated performance management: The evolving South African management model. *Journal of Public Administration*, *41*(1), 177–190.

Modell, S. (2004). Performance measurement myths in the public sector: A research note. *Financial Accountability & Management*, *20*(1), 39–55.

Modipa, M. (2015, March 16). If SA wants to emulate Asia, implementation of NDP is key. *Pretoria News*, p. 14.

Mouton, C. (2010). *The history of programme evaluation in South Africa* (Unpublished MPhil thesis). Stellenbosch University, Stellenbosch.

Mouton, C. (2018). *Performance tracking and the value of data*. PowerPoint presentation presented at meeting with Department of Performance Monitoring and Evaluation (DPME), Johannesburg.

Mouton, J. (2017). *Designing evaluation studies*. PowerPoint presentation presented at the Centre for Research on Evaluation, Science and Technology, Stellenbosch.

Mouton, J. (2019). *Impact Evaluation*. PowerPoint presentation presented at the Centre for Research on Evaluation, Science and Technology, Stellenbosch.

Ndaliso, C. (2019, February 6). Citizens negative about the Future of SA. *Daily News*, p. 1

Neely, A., Adams, C., & Crowe, P. (2001). The performance prism in practice. *Measuring Business Excellence*, *5*(2), 6–12.

Newcomer, K. (1997). Using performance measurement to improve programs. *New Directions for Evaluation*, (75), 5–14.

Nielsen, S., & Hunter, D. E. K. (2013). Challenges to and forms of complementarity between performance management and evaluation. *New Directions for Evaluation*, *137*, 115–123.

Nielsen, S. B., & Ejler, N. (2008). Improving performance? Exploring the complementarities between evaluation and performance management. *Evaluation, 14*(2), 171–192.

Nilsen, P., Stahl, C., Roback, K., & Cairney, P. (2013). Never the twain shall meet? A comparison of implementation science and policy implementation research. *Implementation Science*, *8*(63), 1–12.

Nomsa, N. (N.d.). *South African ICT policy in education.* Sutori. Retrieved from https://www.sutori.com/story/south-african-ict-policy-in-education--4TDZoNikwg7b9U8MpCFi64m4

Noor-Ul-Amin, S. (2013). An effective use of ICT for education and learning by drawing on worldwide knowledge, research, and experience: ICT as a change agent for education. *Scholarly Journal of Education, 2*(4), 38–45.

O'Flynn, J. (2007). From new public management to public value: Paradigmatic change and managerial implications. *Australian Journal of Public Administration*, *66*(3), 353–366.

Organisation for Economic Co-operation and Development (OECD). (2004). "The OECD-JRC Handbook on Practices for Developing Composite Indicators", paper presented at the OECD Committee on Statistics, 7-8 June 2004. Paris: OECD

Organisation for Economic Co-operation and Development (OECD). (2015). *Delivering from the centre. Strengthening the role of the centre of government in driving priority strategies.* Retrieved from https://www.oecd.org/gov/cog-2015-delivering-priority-strategies.pdf

Olson, J., Codde, J., DeMaagd, K., Tarkelson, E., Sinclair, J., Yook, S., & Egidio, R. (2011). *An analysis of eLearning impacts & best practices in developing countries with reference to secondary school education in Tanzania*. East Lansing, MI: Michigan State University.

Osborne, D., & Gaebler, T. (1992). *Reinventing government: How the entrepreneurial spirit is transforming the public sector.* New York, NY: Addison-Wesley.

Osborne, S. P. (2010). Introduction: The new public governance: A suitable case for treatment. In S. P. Osborne (Ed.), *The new public governance? Emerging Perspectives on the Theory and Practice of Public Governance* (pp. 1–13). London: Routledge.

Ospina, S., Grau, N. C., & Zaltsman, A. (2004). Performance evaluation, public management improvement and democratic accountability. *Public Management Review*, *6*(2), 229–251.

O'Toole, L. J. (2004). The theory-practice issue in policy implementation research. *Public Administration*, *82*(2), 309–329.

O'Toole, L. J. (1986). Policy recommendations for multi-actor implementation: An assessment of the field. *Journal of Public Policy*, *6*(2), 181–210.

Owen, J.M., & Rogers, P.J. (1999). Program Evaluation forms and approaches. London: Sage

Palumbo, D. J., & Calista, D. J. (1990a). Introduction: The relation of implementation research to policy outcomes. In D. J. Palumbo & D. J. Calista (Eds.), *Implementation and the policy process: Opening up the black box* (pp. xi–2). Westport, CT: Greenwood Press.

Palumbo, D. J., & Calista, D. J. (1990b). Opening up the black box: Implementation and the policy process. In D. J. Palumbo & D. J. Calista (Eds.), *Implementation and the policy process: Opening up the black box* (pp. 3–18). Westport, CT: Greenwood Press.

Parsons, W. (2002). From muddling through to muddling up: Evidence based policy making and the modernisation of British government. *Public Policy and Administration*, *17*(3), 43–60.

Patton, M. Q. (1978). *Utilization-focused evaluation*. Beverly Hills, CA: Sage.

Peeraer, J., & Van Petegem, P. (2012). Measuring integration of information and communication technology in education: An item response modelling approach. *Computers and Education*, *58*(4), 1247–1259.

Perrin, B. (2002). *Implementing the vision: Addressing challenges to results-focused management and budgeting*. Retrieved from: http://www.oecd.org/governance/budgeting/2497163.pdf

Peters, B. G. (2014). Implementation structures as institutions. *Public Policy and Administration*, *29*(2), 131–144.

Peters, B. G., & Pierre, J. (2001). Developments in intergovernmental relations: Towards multi-level governance. *The Policy Press, 29*(2), 131–135.

Phillips, S., Goldman, I., Gasa, N., Akhalwaya, I., & Leon, B. (2014). A focus on M&E of results: An example from the Presidency, South Africa. *Journal of Development Effectiveness*, *6*(4), 392–406.

Pidd, M. (2005). Perversity in public service performance measurement. *International Journal of Productivity and Performance Management*, *54*(5/6), 482–493.

Poister. T.H. (2010). Performance measurement: monitoring program outcomes. In J.S. Wholey, H.P. Hatry, & K.E. Newcomer (Eds.), *Handbook of practical program evaluation* (3rd ed.) (pp.100-124). San Francisco, CA: Jossey-Bass.

Pollitt, C. (2003). *The essential public manager*. Berkshire, UK: Open University Press.

Pollitt, C., & Bouckaert, G. (2000). *Public management reform: A comparative analysis.* New York, N.Y.: Oxford University Press.

Posner, P. L., & Fantone, D. M. (2007). Assessing federal program performance: Observations on the US Office of Management and Budget's program assessment rating tool and its use in the budget process. *Public Performance & Management Review*, *30*(3), 351–368.

Power, M. (1997). *The audit society: Rituals of verification*. Oxford: Oxford University Press.

Power, M. (2000). The audit society: Second thoughts. *International Journal of Auditing*, *4*(1), 111–119.

Pressman, J., & Wildavsky, A. (1973). *Implementation: How great expectations in Washington are dashed in Oakland; or, why it's amazing that federal programs work at all, this being a saga of the economic development administration as told by two sympathetic observers who seek to build morals on a foundation of ruined hopes*. Berkeley, CA: University of California Press.

Pretoria News Weekend. (2019, January 19). SA policies lack implementation, p. 3.

Price, K. M. (2011). The evolution of understanding: Positioning evaluation within a comprehensive performance management system. *New Directions for Evaluation*, *131*, 103–109.

Radej, B., Golobic, M., & Cernic Istenic, M. (2010). *Beyond new public management doctrine in policy impact evaluation* (Working Paper No. 3[1]). Ljubljana: Slovenian Evaluation Society.

Radnor, Z. J., & Barnes, D. (2007). Historical analysis of performance measurement and management in operations management. *International Journal of Productivity and Performance Management*, *56*(5/6), 384–397.

Rhodes, R. A. W. (1981). *Control and power in central-local government relations.* Farnborough: Gower.

Rhodes, R. A. W. (1988). *Beyond Westminster and Whitehall: The sub-central governments of Britain*. London: Unwin Hyman.

Richards, G., & Chegus, M. (2018). *Does deliverology deliver?* Unpublished paper, Teffler School of Management, University of Ottawa, Ottawa.

Ridgway, V. F. (1956). Dysfunctional consequences of performance measurements. *Administrative Science Quarterly*, *1*(2), 240–247.

Ridley, C. (1927). *Measuring municipal government: Suggested standards for measuring the results of fire, health, police, and public works departments*. New York, NY & Syracuse, NY: New York City Municipal Administration Service & Syracuse University School of Citizenship and Public Affairs.

Ridley, C.E., & Simon, H. (1938). *Measuring municipal activities: A survey of suggested criteria and reporting forms for appraising administration*. Chicago, IL: International City Management Association.

Ridley, C. E., & Simon, H. A. (1943). *Measuring municipal activities: A survey of suggested criteria for appraising administration.* Chicago, IL: The International City Managers' Association.

Rist, R. C. (1989). Management accountability: The signals sent by auditing and evaluation. *Journal of Public Policy*, *9*(3), 355–369.

Rist, R. C. (2006). The "E" in monitoring and evaluation: Using evaluative knowledge to support a results-based management system. In R. C. Rist & N. Stame (Eds.), *From studies to streams: Managing evaluative systems* (pp. 3–22). London: Transaction.

Rogers, P. J. (2008). Using programme theory to evaluate complicated and complex aspects of interventions. *Evaluation*, *14*(1), 29 - 48.

Romzek, B., & Dubnick, M. (1987, May/June). Accountability in the public sector: Lessons from the challenger tragedy. *Public Administration Review, 47,* 227–238.

Roos, M. (2012). Governance and public sector transformation in South Africa. *Africa's Public Service Delivery and Performance Review*, *1*(3), 4–24.

Ropret, M., & Aristovnik, A. (2019). Public sector reform from the post-new public management perspective: Review and bibliometric analysis. *Central European Public Administration Review*, *17*(2), 89–116.

Rosenberg, J.P., & Yates, P.M. (2007). Schematic representation of case study research designs.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA: Sage.

Rossi, R.J. & Gilmartin, K.J. (1980). *The Handbook of social indicators: sources, characteristics, and analysis*. New York, NY: Garland STPM Press

Rowe-Setz, J. (2018). *Western Cape Delivery Support Unit: A case study review*. Unpublished report produced for the Western Cape Delivery Support Unit by Blueprint.

Runya, X., Qigui, S., & Wei, S. (2015). The third wave of public administration: The new public governance. *Canadian Social Science*, *11*(7), 11–21.

Sabatier, P., & Mazmanian, D. (1979). The conditions of effective implementation: A guide to accomplishing policy objectives. *Policy Analysis*, *5*(4), 481–504.

Sabatier, P. A. (1986). Top-down and bottom-up approaches to implementation research: A critical analysis and suggested synthesis. *Journal of Public Policy*, *6*(1), 21–48.

Saeed, M. (2019, July 10). Citizens tired of promises. *Post*, p. 9.

Saetren, H. (2005). Facts and myths about research on public policy implementation: Out-of-fashion, allegedly dead, but still very much alive and relevant. *Policy Studies Journal*, *33*(4), 559–582.

Saetren, H. (2014). Implementing the third generation research paradigm in policy implementation research: An empirical assessment. *Public Policy and Administration*, *29*(2), 84–105.

Sanderson, I. A. N. (2001). Performance management, evaluation and learning in "modern" local government. *Public Administration*, *79*(2), 297–313.

Schacter, M. (2016). *Does deliverology matter?* March Schacter Consulting. Retrieved from https://www.academia.edu/26475943/Does_Deliverology_ Matter

Shaw, J.S. (N.d.). Public Choice Theory. *The Concise Encyclopaedia of Economics*. Retrieved from https://www.econlib.org/library/Enc1/PublicChoiceTheory.html

Schofield, J. (2001). Time for a revival? Public policy implementation: A review of the literature and an agenda for future research. *International Journal of Management Reviews*, *3*(3), 245–263.

Schofield, J., & Sausman, C. (2004). Symposium on implementing public policy: Learning from theory and practice. *Public Administration*, *82*(2), 235–248.

SchoolNet South Africa. (N.d.). *Intel teach project*. Retrieved from https://www.schoolnet.org.za/atwork/intelttf.htm

Schwella, E. (2001). Public sector policy in the new South Africa: A critical review. *Public Performance & Management Review*, *24*(4), 367–388.

Scism, D. (2015). *The new 'deliverology': 14 years of evolution*. London: London School of Economics and Centre for Public Impact. Retrieved from: https://www.slideshare.net/DanielleScism/capstone-report-bcg-final-47696210)

Smith, M. J., Richards, D., Geddes, A., & Mathers, H. (2011). Analysing policy delivery in the United Kingdom: The case of street crime and anti-social behaviour. *Public Administration, 89*(3), 975–1001.

Smith, P. (1990). The use of performance indicators in the public sector. *Journal of the Royal Statistical Society. Series A (Statistics in Society), 153*(1), 53–72.

Smith, P. (1995a). On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration*, 18 (2/3), 277-310.

Smith, P. (1995b). Performance indicators and outcome in the public sector. *Public Money & Management, 15*(4), 13–16.

Smith, S. R., & Smyth, J. (2010). The governance of contracting relationships: "Killing the golden goose" A third-sector perspective. In S. P. Osborne (Ed.), *The new public governance? Emerging Perspectives on the Theory and Practice of Public Governance* (pp. 270–298). London: Routledge.

South African Government. (2019). *Western Cape on enrolment figures for 2019*. Retrieved from: https://www.gov.za/speeches/western-cape-enrolment-figures-2019-1-feb-2019-0000

Starman, A.B. (2013). The case study as a type of qualitative research. *Journal of Contemporary Educational Studies, 1*, 28-43

Sun, P., Tsai, R. J., Finger, G., Chen, Y., & Yeh, D. (2008). What drives a successful eLearning? An empirical investigation of the critical factors influencing learner satisfaction. *Computers and Education, 50*, 1183–1202.

Takyui, E. (2015). The challenge of involvement and detachment in participant observation. *The qualitative report, 20*(6), 864-872.

Theurer, J. (1998). Seven pitfalls to avoid when establishing performance measures. *Public Management, 80*(7), 21–24.

Thomas, G. (2011). A typology for the case study in social science following a review of definition, discourse, and structure. *Qualitative Inquiry, 17*(6), 511-521.

The World Bank. (2017, April). *Driving performance from the center: Malaysia's experience with PEMANDU*. Knowledge & Research: The Malaysia Development Experience Series. Retrieved from https://documents.worldbank.org/en/publication/documents-

reports/documentdetail/318041492513503891/driving-performance-from-the-center-malaysia-s-experience-with-pemandu

Todd, R., Martin, J., & Brock, A. (2014). *Delivery units: Can they catalyse sustained improvements in education service delivery?* Retrieved from https://www.camb-ed.com/intdev/article/252/delivery-units

Toonen, T. (1998, Summer). Networks, management and institutions: Public administration as 'normal science'. *Public Administration, 76*(2), 229–252.

Torfing, J., & Triantafillou, P. (2013). What's in a name? Grasping new public governance as a political-administrative system. *International Review of Public, 18*(2), 9–25.

Umlaw, F., & Chitepo, N. (2015). State and use of monitoring and evaluation systems in national and provincial departments. *African Evaluation Journal*, *3*(1), 1–15.

United Nations Educational, Scientific and Cultural Organization. (2009). *Guide to measuring information and communication technologies, ICT, in education.* (Technical Paper No.2). Retrieved from http://uis.unesco.org/sites/default/files/documents/guide-to-measuring-information-and-communication-technologies-ict-in-education-en_0.pdf

United Nations Programme on HIV/AIDS (UNAIDS). (n.d.). *An introduction to indicators.* Retrieved from https://www.unaids.org/sites/default/files/sub_landing/files/8_2-Intro-to-IndicatorsFMEF.pdf

Vander Elst, S., & De Rynck, F. (2013). Will mandated network steering do the trick? A balanced assessment of the Belgian network "Crossroads Bank for Enterprises". *International Review of Public Administration*, *18*(2), 47–64.

Van der Heijden, T., & Mlandi, M. (2005, October). Organisational performance, success and failure in the public sector. *Journal of Public Administration, 10*(1), 234–247.

Van der Waldt, G. (2006). Managing local government performance: Key considerations and challenges. *Journal of Public Administration*, *41*(2), 128–143.

Van Meter, D. S., & Van Horn, C. E. (1975). The policy implementation process: A conceptual framework. *Administration & Society*, *6*(4), 445–487.

Verschuere, B. (2009). The role of public agencies in the policy making process: Rhetoric versus reality. *Public Policy and Administration, 24*(1), 23–46.

Voets, J., Van Dooren, W., & De Rynck, F. (2008). A framework for assessing the performance of policy networks. *Public Management Review*, *10*(6), 773–790.

Wahl, D.C. (2017). Facing complexity: Wicked design problem. Excerpt from 2006 PhD thesis: Design for human and planetary health. A holistic/ integrated approach to complexity and sustainability. (Unpublished doctoral dissertation). University of Dundee, Dundee.

Weiss, C. H. (Ed.). (1977). *Using social research in public policy-making*. Lexington, MA: Lexington Books.

Weiss, C.H. (1998). *Evaluation: methods for studying programs and policies*. (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Western Cape Education Department. (2012). *Vision for e-Education: eLearning and e-Teaching in schools of the future*. PowerPoint presentation presented at WCED Top Management meeting, Cape Town.

Western Cape Education Department. (2015). *E-education plan: 2015/16–2019/2020*. Western Cape Education Department, Cape Town

Western Cape Government. (2014). *2014 State of the Province Address by Premier Helen Zille*. Retrieved from https://www.westerncape.gov.za/speech/2014-state-province-address-Premier-helen-zille

Western Cape Government, Department of the Premier. (2019). *Annual performance plan 2019/20*. Retrieved from https://www.westerncape.gov.za/sites/www.westerncape.gov.za/files/assets/departments/Premier/dotp_app2019_final.pdf

Western Cape Government, Delivery Support Unit. (2016a). *Game Changer roadmap 2016 eLearning*. Internal Western Cape Government report: unpublished report

Western Cape Government, Delivery Support Unit. (2016b). *eLearning stocktake report 15 March 2016*. [Powerpoint presentation]

Western Cape Government, Delivery Support Unit. (2017a). *eLearning stocktake report 7 February 2017*. [Powerpoint presentation]

Western Cape Government, Delivery Support Unit. (2017b). *eLearning stocktake report 6 June 2017.* [Powerpoint presentation]

Western Cape Government, Delivery Support Unit. (2018a). *eLearning stocktake report 20 February 2018.* [Powerpoint presentation]

Western Cape Government, Delivery Support Unit. (2019a). *Handover issues of the eLearning Game Changer*. Unpublished internal report, J.C. Cargill, C. Mouton and P. Tainton, Cape Town.

Western Cape Government, Delivery Support Unit. (2019b). *eLearning stocktake report 28 February 2019*. [Powerpoint presentation]

Western Cape Government, Delivery Support Unit. (2019c). *eLearning stocktake report 16 April 2019.* [Powerpoint presentation]

Wholey, J. S., & Hatry, H. P. (1992). The case for performance monitoring. *Public Administration Review, 52*(6), 604–610.

Wholey, J. S., Hatry, H. P., & Newcomer, K. E. (Eds.) (2010). *Handbook of practical program evaluation* (3rd ed.). San Francisco, CA: Jossey-Bass.

Wholey, J. S., & Newcomer, K. E. (1997, Fall). Clarifying goals, reporting results. *New Directions, 75*, 91–98.

Wildschut, L. (2019). *Understanding programmes*. PowerPoint presentation presented at the Centre for Research on Evaluation, Science and Technology, Stellenbosch.

Willems, T., & Van Dooren, W. (2011). Lost in diffusion? How collaborative arrangements lead to an accountability paradox. *International Review of Administrative Sciences, 77*(3), 505–530.

Williams, D. W. (2002). Before performance measurement. *Administrative Theory & Praxis, 24*(3), 457–486.

Williams, D. W. (2003). Measuring government in the early twentieth century. *Public Administration Review, 63*(6), 643–659.

Williams, D. W. (2004). Evolution of performance measurement until 1930. *Administration & Society, 36*(2), 131–165.

Winter, S. (1990). Integrating implementation research. In D. J. Palumbo & D. J. Calista (Eds.), *Implementation and the policy process: Opening up the black box* (pp. 19–38). Westport, CT: Greenwood Press.

Winter, S. C. (2003). Implementation perspectives. In B. Guy Peters & J. Pierre (Eds.), *Handbook of public administration* (pp. 212–223). London: Sage.

Yin, R.K. (2009). *Case study research: design and methods*. (4th ed.) Los Angeles, CA: Sage

Zumofen, R. (2016). *Redefining accountability in a strategic perspective to enhance performance*. Paper presented to the 2015 International Research Society for Public Management (IRSPM) Conference, University of Birmingham, Birmingham.